



TRABAJO FIN DE GRADO  
GRADO DE INGENIERÍA DEL SOFTWARE  
CURSO 2015-2016

---

**Identificación de la Fuente en Ficheros  
Multimedia de Redes Sociales**

**Guillermo Rodríguez Franco**

Directores:

**Luis Javier García Villalba**

**Ana Lucila Sandoval Orozco**

Departamento de Ingeniería del Software e Inteligencia Artificial

---

FACULTAD DE INFORMÁTICA  
UNIVERSIDAD COMPLUTENSE DE MADRID



El presente Trabajo Fin de Grado se enmarca dentro de un proyecto de investigación titulado RAMSES aprobado por la Comisión Europea dentro del Programa Marco de Investigación e Innovación Horizonte 2020 y en el que participa el Grupo GASS del Departamento de Ingeniería del Software e Inteligencia Artificial de la Facultad de Informática de la Universidad Complutense de Madrid (Grupo de Análisis, Seguridad y Sistemas, <http://gass.ucm.es>, grupo 910623 del catálogo de grupos de investigación reconocidos por la UCM).

Por razones de confidencialidad del proyecto se ha omitido información del trabajo desarrollado para no infringir la normativa correspondiente.

---

Guillermo Rodríguez Franco

---

Luis Javier García Villalba

---

Ana Lucila Sandoval Orozco



## *Agradecimientos*

A mis directores Luis Javier García Villalba y Ana Lucila Sandoval Orozco por haberme dado la oportunidad de trabajar con ellos, por haber hecho posible la realización de este trabajo fin de grado y por toda su dedicación durante estos meses de duro trabajo.

A todos mis compañeros de la Facultad, que he conocido a lo largo de todos estos años, por vuestra amistad, vuestros ánimos y vuestro apoyo, en especial a mi compañera Lidia María Jiménez Gutiérrez.

A todos los compañeros y ex compañeros del trabajo por su apoyo y ayuda en todo momento.

Un especial agradecimiento para todo mi grupo de amigos del Colegio Nuestra Señora del Pilar, por haber estado ahí en todos mis momentos difíciles y haberme apoyado siempre en absolutamente todo.

Mi más emotivo y sentimental agradecimiento está dedicado a mi familia, a mis padres y a mi compañera de viaje Rebeca Pérez Moreno, por todo su apoyo, cariño y esfuerzo, que me han permitido realizar todos mis estudios.



## *Resumen*

El impacto en la vida cotidiana que tiene el uso de dispositivos móviles para capturar imágenes digitales alimentó un creciente interés en compartir fotos y vídeos a través de las redes sociales. Desde un punto de vista forense, los usuarios de Internet, y en especial de estas redes sociales, dejan continuamente rastros visuales de su presencia en Internet, ofreciendo una valiosa información útil para las investigaciones.

Las herramientas de análisis forense de imágenes digitales permiten recolectar evidencias de estas imágenes. Una de estas técnicas utiliza el patrón del ruido del sensor (SPN) que actúa como una huella digital en la imagen del sensor de la cámara que la generó.

En este trabajo se propone utilizar el patrón PRNU, combinado con técnicas de agrupamiento, para identificar al propietario de un dispositivo móvil, analizando las fotos que un usuario de redes sociales sube desde el mismo dispositivo. Relacionar una imagen con el propietario de un dispositivo puede ser muy útil en una variedad de casos forenses: *ciberbullying*, abuso infantil en línea, búsqueda de cámaras robadas, etc.

La evaluación de la propuesta presentada se realizó con fotos subidas a 3 redes sociales por usuarios creados para este fin. Los resultados obtenidos presentan una tasa de acierto elevada, y demuestran el beneficio de usarla como herramienta forense.

## *Palabras clave*

Adquisición de imágenes, agrupamiento, análisis forense, cámara de dispositivos móviles, redes sociales, PRNU.





## *Abstract*

The impact on everyday life that has the use of mobile devices to capture digital images fueled a growing interest in sharing images and videos through social networks. From a forensic point of view, Internet users, and especially of these social networks, continuously leave visual traces of their presence on the Internet, providing valuable information useful for investigations.

Forensic analysis tools allow digital imaging collect evidence of these images. One of these techniques uses the pattern sensor noise (SPN) that acts as a fingerprint image of the camera sensor that generated it.

This paper proposes to use the PRNU pattern, combined with clustering techniques to identify the owner of a mobile device, analyzing the photos that a user of social networks uploads from the same device. Relate an image with the device owner can be very useful in a variety of forensic cases: cyberbullying, online child abuse, search for stolen cameras, etc.

The evaluation of the proposal was made with pictures uploaded by users (created for this purpose) to 3 social networks. The results obtained show a high rate of success, and demonstrate the benefit of using it as a forensic tool.

## *Keywords*

Image acquisition, camera mobile devices, clustering, forensics analysis, social networks, PRNU.



## ***Lista de Acrónimos***

API	Application Programming Interface
CFA	Color Filter Array
GNU	General Public License
GPS	Global Positioning System
HTML	HyperText Markup Language
IQM	Image Quality Metrics
JPEG	Joint Photographic Experts Group
LECr	Ley de Enjuiciamiento Criminal
OMS	Organización Mundial de la Salud
PRNU	Photo Response Non Uniformity
SPN	Sensor Pattern Noise
SVM	Support Vector Machine
TPR	True Positive Rate
URL	Uniform Resource Identifier



# ÍNDICE

<b>1. INTRODUCCIÓN .....</b>	<b>1</b>
1.1. MOTIVACIÓN .....	1
1.2. CONTEXTO .....	4
1.3. OBJETO DEL PROYECTO .....	5
1.4. PLANIFICACIÓN DEL TRABAJO .....	5
1.5. ESTRUCTURA DE LA MEMORIA .....	9
<b>2. USO DE REDES SOCIALES DESDE DISPOSITIVOS MÓVILES.....</b>	<b>11</b>
2.1. DEFINICIÓN DE RED SOCIAL .....	11
2.2. USOS DE UNA RED SOCIAL EN LA SOCIEDAD ACTUAL .....	12
2.3. TIPOS DE REDES SOCIALES .....	13
2.4. REDES SOCIALES MÁS POPULARES .....	14
2.5. ADICCIÓN .....	16
2.5.1. ADICCIÓN AL TELÉFONO MÓVIL .....	16
2.5.2. ADICCIÓN A LAS REDES SOCIALES .....	17
2.6. USO DE IMÁGENES Y VÍDEOS EN LAS REDES SOCIALES .....	18
<b>3. ESTADO DEL ARTE.....</b>	<b>21</b>
3.1. TÉCNICAS DE IDENTIFICACIÓN DE LA FUENTE .....	22
3.2. TÉCNICAS DE CLUSTERING .....	24
<b>4. THEIA: HERRAMIENTA PARA EL ANÁLISIS FORENSE DE IMÁGENES Y VÍDEOS DIGITALES.....</b>	<b>27</b>
4.1. TRATAMIENTO A NIVEL INDIVIDUAL .....	27
4.2. TRATAMIENTO A NIVEL DE GRUPO .....	30
<b>5. CONTRIBUCIÓN .....</b>	<b>35</b>
5.1. CONSIDERACIONES GENERALES .....	35
5.2. FUNCIONAMIENTO .....	36
5.3. HERRAMIENTAS DE DESARROLLO.....	39
5.4. EVALUACIÓN DE LOS ALGORITMOS .....	40
5.4.1. CONFIGURACIÓN DE LOS EXPERIMENTOS .....	40
5.4.2. EXPERIMENTOS .....	43
<b>6. CONCLUSIONES Y TRABAJO FUTURO.....</b>	<b>45</b>
6.1. CONCLUSIONES.....	45
6.2. TRABAJO FUTURO.....	46
<b>RESUMEN EN INGLÉS</b>	
<b>7. INTRODUCTION .....</b>	<b>49</b>
7.1. MOTIVATION.....	49
7.2. PROJECT GOAL .....	51
7.3. WORK PLANIFICATION.....	52
7.4. PROJECT STRUCTURE.....	55
<b>8. CONCLUSIONS AND FUTURE WORK .....</b>	<b>57</b>
8.1. CONCLUSIONS.....	57
8.2. FUTURE WORK .....	58
<b>REFERENCIAS .....</b>	<b>59</b>

## ÍNDICE DE TABLAS

Tabla 1.1. Fases del proyecto .....	6
Tabla 1.2. Actividades de la fase de ejecución del proyecto.....	7
Tabla 2.1. Tipos de redes verticales.....	13
Tabla 4.1. Dispositivos móviles utilizados en los experimentos.....	41
Tabla 4.2. Resoluciones de descarga disponibles en las redes sociales.....	42
Tabla 4.3. Matriz de Confusión del experimento 1 con una tasa de acierto del 82.8% ...	44
Tabla 4.4. Matriz de Confusión del experimento 2 con una tasa de acierto del 88.8% ...	44
Tabla 6.1. Project phases.....	53

## ÍNDICE DE FIGURAS

Figura 1.1: Incremento de usuarios de teléfonos móviles y redes sociales entre 2014 y 2019.....	2
Figura 1.2: Diagrama de Gantt de la planificación del proyecto .....	8
Figura 2.1 Redes sociales más importantes .....	14
Figura 4.1. Apariencia general de la pestaña <i>Exif Info</i> .....	27
Figura 4.2. Geoposicionamiento en Google Maps .....	29
Figura 4.3. Apariencia general de la pestaña <i>DDBB Projects</i> .....	30
Figura 4.4. Visualización de las imágenes de un proyecto.....	31
Figura 4.5. Query Set.....	32
Figura 4.6. <i>Advanced Query</i> .....	33
Figura 4.7. Geoposicionamiento de un grupo de imágenes en Google Maps .....	34
Figura 5.1 Esquema general de funcionamiento del algoritmo.....	37
Figure 6.1: Mobile devices and social networks users between 2014 and 2019 .....	50





# 1. INTRODUCCIÓN

---

## 1.1. Motivación

Desde que apareció el primer teléfono móvil, creado por Martin Cooper en 1973, se han producido cambios significativos en el diseño, la tecnología y el uso del mismo. Ha pasado a ser un medio que ofrece diversidad de servicios integrados como: una cámara cada vez más eficiente y mejorada, una mayor capacidad de almacenamiento interno y servicios a través de la conexión a internet, por ejemplo, acceso a las redes sociales, comunicación por mensajería instantánea con aplicaciones como WhatsApp, Skype, Telegram o Facebook.

Todo ello ha llevado a una dependencia del teléfono móvil, teniendo a los usuarios localizados y pendientes del mismo en todo momento. El móvil se ha convertido en un instrumento básico y necesario en la vida de cualquier adolescente y adulto, sirviéndose de éstos para sentirse integrados en la sociedad, interactuando entre su grupo de iguales y estando alerta de las últimas novedades. Asimismo, puede provocar una frustración por el hecho de no estar al día y a la moda. Como se observa, el teléfono móvil forma parte del día a día de la sociedad actual, sin el que difícilmente se podría sobrevivir hoy en día. Esto lo demuestran estudios realizados por Statista Inc. [1][2]. En la Figura 1.1 se muestran los resultados de dichos estudios. Como se observa en la figura, en los últimos años hay un incremento constante en el número de usuarios con teléfonos móviles y el número de usuarios registrados en redes sociales. Este incremento ha sido similar en ambos casos y continuará así hasta el año 2019.

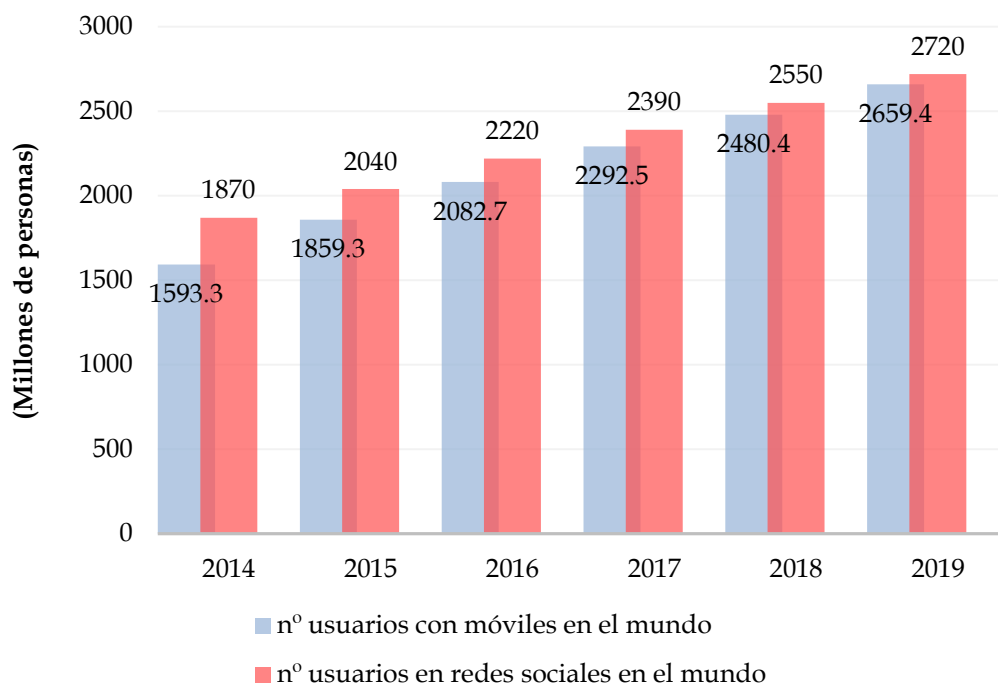


Figura 1.1: Incremento de usuarios de teléfonos móviles y redes sociales entre 2014 y 2019

Como consecuencia de lo anterior, cada vez es más común presentar como prueba en un juicio, mensajes de WhatsApp, SMS, fotos, vídeos o comentarios compartidos en las redes sociales como indicios o pruebas de la investigación de un delito. Estas pruebas son incorporadas a un caso en papel, como prueba documental, la cual deben ser obtenidas con plenas garantías de legalidad, de tal modo que quien proporcione dichas pruebas debe tener derecho para la posesión y uso de las mismas, por ser el receptor o el emisor de las mismas.

En el procedimiento penal español no existe una medida específica sobre el tratamiento de pruebas de tales comunicaciones, por tanto, se ha de acudir a los siguientes artículos:

- **Artículo 26 del Código Penal:** Establece que “A los efectos de este código se considera documento todo soporte material que exprese o incorpore datos, hechos o narraciones con eficacia probatoria o cualquier otro tipo de relevancia jurídica” [3].

- **Artículo 726 LECr:** Establece que “El Tribunal examinará por sí mismo los libros, documentos, papeles y demás piezas de convicción que puedan contribuir al esclarecimiento de los hechos o a la más segura investigación de la verdad” [4].

Los juzgados y tribunales suelen admitir dichas pruebas e incorporarlas al procedimiento luego de realizar una comprobación de las mismas. Tiene que quedar constancia la marca, el modelo del dispositivo del que proceden las pruebas.

A continuación, se muestran algunos casos en los que parte de las pruebas involucran teléfonos móviles:

En Estados Unidos se investigó en el año 2015, en la localidad de Jeanette (Pensilvania), un caso de asesinato que giró en torno a un mensaje de Snapchat, que se usó como principal prueba del caso. El caso trataba sobre una adolescente que se enfrentaba a una acusación de asesinato en primer grado de otro joven, donde la principal prueba era una captura de pantalla de un mensaje de Snapchat, cuyo contenido era una foto de la presunta asesina con el cadáver de la víctima [5]. Un caso similar, fue el de una persona que publicó una foto suya junto al cadáver de su pareja en Facebook después de asesinarla [6]. Finalmente, en 2014, una persona fue detenida en Washington D.C. acusada del asesinato de su pareja tras obtener unas fotos publicadas en el foro 4chan del cadáver de la víctima [7]. En estos casos se utilizó el rastro dejado por los sospechosos en Internet como prueba de su culpabilidad.

Teniendo en cuenta el uso abusivo tanto del teléfono móvil como de las redes sociales y todo lo anterior relacionado con pruebas para la resolución de un caso, se hace necesario llevar a cabo una investigación de todo lo relacionado con usuarios altamente sospechosos de haber llevado a cabo alguna incoherencia, ya que cualquier pista, imagen o amenaza puede ayudar a lograr identificar la fuente que generó la imagen o el vídeo presentado como prueba.

## 1.2. Contexto

El presente Trabajo Fin de Grado se enmarca dentro de un proyecto de investigación titulado RAMSES aprobado por la Comisión Europea dentro del Programa Marco de Investigación e Innovación Horizonte 2020 (Convocatoria H2020-FCT-2015, Acción de Innovación, Número de Propuesta: 700326) y en el que participa el Grupo GASS del Departamento de Ingeniería del Software e Inteligencia Artificial de la Facultad de Informática de la Universidad Complutense de Madrid (Grupo de Análisis, Seguridad y Sistemas, <http://gass.ucm.es>, grupo 910623 del catálogo de grupos de investigación reconocidos por la UCM).

Además de la Universidad Complutense de Madrid participan las siguientes entidades:

- Treeologic Telemática y Lógica Racional para la Empresa Europea SL (España)
- Ministério da Justiça (Portugal)
- University of Kent (Reino Unido)
- Centro Ricerche e Studi su Sicurezza e Criminalità (Italia)
- Fachhochschule für Öffentliche Verwaltung und Rechtspflege in Bayern (Alemania)
- Trilateral Research & Consulting LLP (Reino Unido)
- Politecnico di Milano (Italia)
- Service Public Fédéral Intérieur (Bélgica)
- Universität des Saarlandes (Alemania)
- Dirección General de Policía - Ministerio del Interior (España)

### **1.3. Objeto del Proyecto**

La finalidad de este proyecto es analizar y comprender el uso de las redes sociales en los dispositivos móviles y, cómo son las subidas y descargas de las imágenes en las redes sociales realizadas con los dispositivos móviles. Posteriormente, diseñar una herramienta capaz de descargar las imágenes de estas redes sociales, para luego realizar la identificación de la fuente de adquisición de dichas imágenes. Para ello, los objetivos específicos del proyecto son los siguientes:

- Comprender el concepto de red social: En qué consisten, cómo se utilizan, qué tipos hay, la subida y descarga de imágenes en ellas, cuál es el tratamiento dado a las imágenes.
- Analizar las investigaciones realizadas sobre los procesos realizados con imágenes para el análisis forense. Más concretamente, sobre los procesos de identificación de la fuente de adquisición de imágenes.
- Desarrollar una herramienta capaz de descargar las imágenes de un álbum de un usuario registrado en una red social.
- Implementar un algoritmo de identificación de la fuente imágenes descargadas de las redes sociales.
- Probar la herramienta generada con usuarios de Facebook y Flickr para verificar que funcionalidad y fiabilidad.

### **1.4. Planificación del Trabajo**

El proyecto se ha desarrollado en 3 fases: Planificación, Ejecución y Documentación del Proyecto. Las actividades realizadas en cada una de estas fases se presentan en la Tabla 1.1.

Nombre de tarea	Tiempo (días)	Inicio	Fin
<b>Planificación del proyecto</b>	<b>35</b>	<b>02/11/15</b>	<b>18/12/15</b>
• Reuniones semanales para y definición con los tutores	35	02/11/15	18/12/15
• Estudio de las técnicas forenses de identificación de la fuente de imágenes de dispositivos móviles	15	02/11/15	20/11/15
• Estudio de las diferentes redes sociales y sus características	13	23/11/15	09/12/15
• Definición del proyecto	7	10/12/15	18/12/15
<b>Ejecución del Proyecto</b>	<b>140</b>	<b>25/01/16</b>	<b>05/08/16</b>
• Especificación de requisitos	55	25/01/16	08/04/16
• Diseño	30	11/04/16	20/05/16
• Implementación	29	23/05/16	30/06/16
• Pruebas	62	20/04/16	14/07/16
• Control	134	11/01/16	14/07/16
<b>Documentación</b>	<b>200</b>	<b>02/11/15</b>	<b>05/08/16</b>
• Generación de documentación del proyecto	184	02/11/15	14/07/16
• Preparación y revisión de la memoria	26	01/07/16	05/08/16

Tabla 1.1. Fases del proyecto

En la fase de ejecución se realiza el desarrollo del proyecto definido en la fase de planificación. En ella se realizaron las siguientes actividades generales: Identificación de requisitos, diseño, construcción y pruebas. Asimismo, se realizaron actividades de seguimiento y control del avance del proyecto, con el objetivo de agilizar los ajustes necesarios en cada una de las actividades. La Tabla 1.2 presenta las actividades realizadas en esta fase, así como la duración de las mismas. La Figura 1.2 muestra el diagrama de Gantt del proyecto.

Nombre de tarea	Tiempo (días)	Inicio	Fin
<b>Especificación de requisitos</b>	<b>55</b>	<b>25/01/16</b>	<b>08/04/16</b>
• Estudio del uso de las diferentes redes sociales en dispositivos móviles	10	25/01/16	05/02/16
• Análisis del proceso de carga y descarga de fotos y vídeos a diferentes redes sociales	10	08/02/16	19/02/16
• Análisis de las transformaciones realizadas a las imágenes y vídeos de las redes sociales	10	22/02/16	04/03/16
• Estudio de técnicas de procesamiento de imágenes	10	07/03/16	18/03/16
• Identificación de las redes sociales orientadas a la compartición de imágenes y vídeos	5	21/03/16	25/03/16
• Análisis del funcionamiento de las redes sociales Facebook, Flickr, Google +, Tuenti, Twitter e Instagram	10	28/03/16	08/04/16
<b>Diseño</b>	<b>30</b>	<b>11/04/16</b>	<b>20/05/16</b>
• Análisis del proceso de Subida y descarga manual de fotos y vídeos a las redes sociales Facebook, Flickr, Google +, Tuenti, Twitter e Instagram	7	11/04/16	19/04/16
• Diseño del algoritmo de descarga de fotos de Facebook	7	20/04/16	28/04/16
• Diseño del algoritmo de descarga de fotos de Flickr	7	29/04/16	09/05/16
• Diseño del módulo de gestión de descargas	2	10/05/16	11/05/16
• Búsqueda y análisis de paquetes o clases que se comuniquen con las redes sociales de Facebook y Flickr	7	12/05/16	20/05/16
<b>Implementación</b>	<b>29</b>	<b>23/05/16</b>	<b>30/06/16</b>
• Estudio de API de programación en Facebook	5	23/05/16	27/05/16
• Implementación del algoritmo de descarga de fotos de Facebook	7	30/05/16	07/06/16
• Estudio de API de programación en Flickr	5	08/06/16	14/06/16
• Implementación del algoritmo de descarga de fotos de Flickr	6	15/06/16	22/06/16
• Implementación de los ajustes del algoritmo de identificación de la fuente de adquisición	6	23/06/16	30/06/16
<b>Pruebas</b>	<b>62</b>	<b>20/04/16</b>	<b>14/07/16</b>
• Creación del dataset de entrenamiento con fotos y vídeos tomados de dispositivos móviles de diferentes marcas y modelos	25	20/04/16	24/05/16
• Algoritmo de descarga de fotos de Facebook	3	24/06/16	28/06/16
• Algoritmo de descarga de fotos de Flickr	3	01/07/16	05/07/16
• Proceso de identificación de fotos de dispositivos móviles descargadas de las redes sociales Facebook y Flickr	7	06/07/16	14/07/16
<b>Control</b>	<b>134</b>	<b>11/01/16</b>	<b>14/07/16</b>
• Reuniones semanales de seguimiento	134	11/01/16	14/07/16

Tabla 1.2. Actividades de la fase de ejecución del proyecto

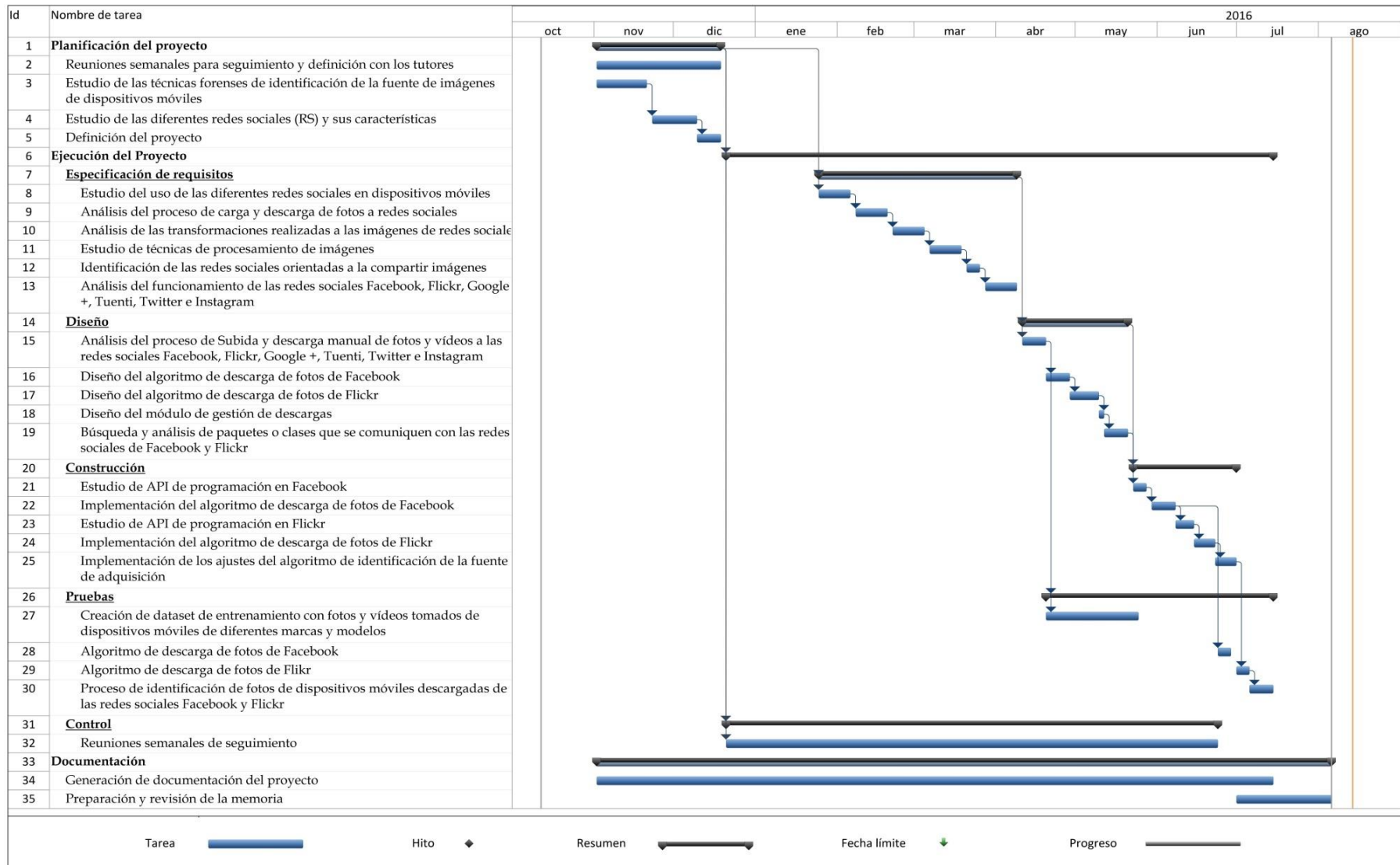


Figura 1.2: Diagrama de Gantt de la planificación del proyecto



## **1.5. Estructura de la Memoria**

En este capítulo y a modo de introducción, se ha descrito la motivación que ha llevado a la realización de este proyecto y el objeto del mismo.

En el capítulo 2, se habla sobre las redes sociales, su uso y clasificación, detallando las más importantes. Asimismo, se menciona la adicción que hoy en día tiene la sociedad con las redes sociales y con los teléfonos móviles, ya que el uso de vídeos e imágenes en las redes sociales se ha incrementado notablemente.

En el capítulo 3 se identifican las diferentes técnicas existentes para el tratamiento e identificación de la fuente de imágenes de dispositivos móviles, y el uso de las imperfecciones del sensor

En el capítulo 4 se presenta la contribución de este trabajo. En él se detallan los algoritmos de descarga de imágenes de las redes sociales Facebook y Flickr, y el de identificación de la fuente de imágenes de las redes sociales. Asimismo, se presenta las pruebas realizadas para evaluar el funcionamiento de los algoritmos.

Por último, en el capítulo 5 se presentan las conclusiones a las que se ha podido llegar en este trabajo, y el trabajo futuro que se desprende del mismo.



## **2. USO DE REDES SOCIALES DESDE DISPOSITIVOS MÓVILES**

---

El objetivo de este capítulo es entender la evolución e importancia de las redes sociales en los últimos años en la sociedad actual. Asimismo, se presenta información sobre el riesgo de adicción que representa el uso de las redes sociales y dispositivos móviles.

### **2.1. Definición de Red Social**

En [8] se define una red social como «un conjunto de lazos entre diversos actores, todos del mismo o similar tipo. Se puede comparar a un club en el que los socios se encuentran, se saludan, intercambian impresiones, ideas, noticias, interactúan en diferentes aspectos y se prestan los unos a los otros, ampliando cada uno su red de amistades, conocimientos y contactos». Una red social tiene un objetivo y unas normas como todos los grupos. Asimismo, el objetivo de un grupo no es la suma de los objetivos de sus componentes, sino que el grupo se conforma para marchar en un sentido, cuenta con unos recursos y admite miembros que se interesen por seguir ese sentido de marcha.

Este concepto ha ido ganando mayor importancia en estos últimos 10 años con la creación de Facebook, Twitter e Instagram (entre otros), convirtiéndose en una forma de expresión, relación y comunicación de las personas. Las redes sociales siempre han sido objeto de estudios sobre todo en antropología y la sociología [9][10], donde se han formado diferentes tipos de teorías para poder explicar el funcionamiento y el porqué de su desarrollo virtual [11]. Las redes sociales han ocupado el primer lugar en Internet, siendo este un campo importante en las relaciones personales e interpersonales. Las primeras redes sociales se crearon en Estados Unidos entre los años 2001- 2003, redes sociales como MySpace o Tribe, que dieron comienzo a las primeras interacciones e intercambio de información entre usuarios [12].

## 2.2. Usos de una Red Social en la Sociedad Actual

Las primeras redes sociales tenían como objetivo relacionar usuarios, buscar y ser buscado por amigos, familiares y conocidos. Actualmente, las redes sociales permiten las siguientes funciones:

- **Mantener la comunicación con la gente que se conoce:** la principal función es formar comunidades.
- **Encontrar información:** hoy en día un individuo puede estar actualizado de lo que sucede en el mundo a través de las redes sociales antes de salir la información en los medios de comunicación, como es el caso de Facebook o Twitter. También se puede encontrar información haciendo una pregunta en las redes sociales para obtener la respuesta y/o saber la opinión del resto de usuarios.
- **Autopromoción:** gracias a las redes sociales, un usuario se puede dar a conocer, promocionarse, compartir su trabajo, su creación...
- **Networking:** algunos usuarios prefieren las redes sociales para el ámbito profesional. Son lugares ideales para las relaciones interpersonales como es el caso de LinkedIn, donde no sólo se trata de conocer a gente con tus mismos fines, sino que puedes establecer y construir conexiones para tener oportunidades de empleo o de negocios.
- **Construir marca:** usar las redes sociales para fortalecer la reputación de una empresa, generar credibilidad y confianza. La imagen de una empresa puede parecer frágil en las redes sociales porque está expuesta a los comentarios directos de los usuarios y los clientes, pero hay que considerar que ninguna otra herramienta da tanto poder a las personas frente a las empresas.

## 2.3. Tipos de Redes Sociales

Se pueden diferenciar dos grandes grupos de redes sociales, las redes horizontales y las redes verticales, a continuación, se explicará brevemente un poco de cada una de ellas.

- **Redes Horizontales:** Las redes sociales horizontales no están orientadas a un público en concreto, sino a todo el mundo, centrándose en los contactos. No tienen un tema específico. La función principal de las redes es relacionar personas a través de las herramientas que ofrecen. Todas comparten las mismas características: crear un perfil, compartir contenidos y generar listas de contactos. Algunas de ellas son: Facebook, Hi5, MySpace, Tuenti, Google+, Badoo. Lo que motiva a los usuarios pertenecientes a estas redes es la interrelación general, sin un propósito concreto. [13]
- **Redes Verticales:** En estos últimos años han aparecido una gran variedad de redes sociales especializadas, las redes sociales verticales las clasifica según su temática, su actividad o por su contenido compartido. Se crean para dar cabida a los gustos e intereses de las personas que buscan un espacio de intercambio común. Por tanto, las redes sociales verticales se caracterizan por la especialización [13]. La Tabla 2.1. presenta la clasificación de este tipo de redes sociales.

Tipo	Especialización
Temática	Profesionales, aficiones, movimientos sociales, viajes.
Actividad	Microblogging (redes sociales que ofrecen un servicio de envío y publicación de mensajes breves de texto), juegos, geolocalización, marcadores sociales, compartir objetos.
Contenido compartido	Fotos, música, vídeos, documentos, presentaciones, lectura, noticias.

Tabla 2.1. Tipos de redes verticales

## 2.4. Redes Sociales más Populares

El uso de las redes sociales se ha incrementado muy notablemente. La Figura 2.1 presenta las redes sociales más importantes. A continuación, se describen cada una de ellas en orden de popularidad.

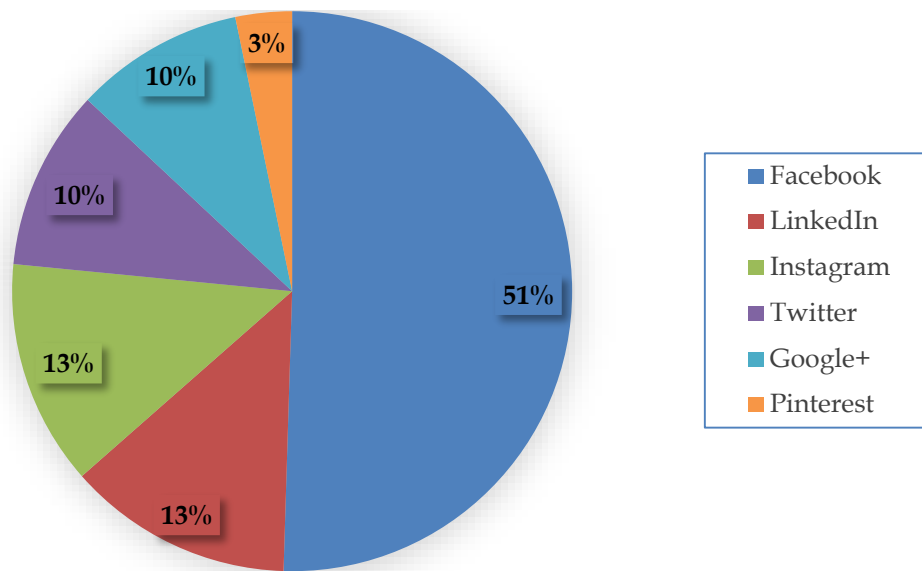


Figura 2.1 Redes sociales más importantes

- **Facebook [14]:** Empezó siendo una red social para estudiantes universitarios. Las relaciones interpersonales siguen siendo la principal razón para unirse a la comunidad. Es la red social más utilizada por los internautas, actualmente cuenta con más de 1.100 millones de usuarios en el mundo, exceptuando China. Está destinada a todo tipo de público, pero sus usuarios están comprendidos entre los 18 años y 50 años aproximadamente, aunque el grupo de edad más activo es de 30 a 44 años. Facebook es un espacio personal donde se puede conectar con los amigos, conocidos y compartir cualquier cosa con ellos, además de crear grupos, hacer video llamadas, jugar a juegos, etc.
- **YouTube [15]:** Es la segunda red social más utilizada a día de hoy con un millón de usuarios. Tiene como objetivo subir vídeos a Internet Los vídeos

subidos se pueden ver, comentar y votar. Asimismo, se pueden subir vídeos en propiedad. En el 2006 esta red fue adquirida por Google, que manifestó su gran potencial, pues es la red más utilizada para subir vídeos y publicarlos. Red social empleada por muchos artistas hoy en día para promocionarse y darse a conocer a la sociedad.

- **Twitter [16]:** Nació en 2006 y ya cuenta con más de 500 millones de usuarios. Es una red de microbloggings fácil de usar y es la más usada porque se puede estar al día de lo que pasa en el mundo, de los amigos, de los ídolos favoritos, etc. Todo lo que se ponga en Twitter se hará viral al segundo.
- **Google+ [17]:** Cuenta con más de 300 millones de usuarios y pertenece a la red social de Google. Esta red es utilizada para crear eventos y círculos entre las personas. Lo bueno de esta red es que se pueden compaginar todas las herramientas que ofrece Google y así tenerlo todo sincronizado a una cuenta propia.
- **LinkedIn [18]:** Surgió en el 2003. Actualmente cuenta con más de 260 millones de usuarios y es la red más usada entre profesionales. Esta red se centra principalmente en buscar trabajo, donde además los contactos pueden valorar las habilidades y capacidades que incluya un usuario en su descripción y CV. También ayuda a los usuarios a generar negocios, mantener contacto directo con líderes del mismo sector, encontrar y ser encontrado por profesionales, obtener respuestas sobre un tema en concreto, intercambiar opiniones, encontrar clientes, proveedores y empleados.
- **Instagram [19]:** Fue creada por Kevin Systrom y Mike Krieger y fue lanzada en octubre de 2010, es la red que mayor crecimiento ha tenido en tan poco tiempo con más de 150 millones de usuarios en la red. El rango de edad de esta red social está entre los 13 años en adelante. Se caracteriza por subir fotos y vídeos al momento a cualquier hora del día, demostrando la creatividad y originalidad de cada uno. Es una red donde sus usuarios

pueden aplicar efectos fotográficos antes de compartir sus fotos.

- **Pinterest [20]:** Cuenta con más de 70 millones de usuarios. Esta red sirve para compartir imágenes de cualquier sitio, donde el usuario puede gestionar sus propios tableros para diferenciarlos de otras categorías de imágenes. Así mismo, el usuario puede seguir otros tableros de los usuarios para mayor creatividad.
- **Flickr [21]:** Fue creada en el 2004 y cuenta con más de 92 millones de usuarios registrados para almacenar, ordenar, buscar, vender y compartir fotografías o vídeos a través de Internet. Cuenta con una comunidad de usuarios que comparten fotografías y vídeos creados por ellos mismos. La principal característica de esta red es la capacidad para administrar imágenes mediante herramientas que permiten a los autores etiquetar sus fotografías, explorar y comentar las imágenes de otros usuarios.

## **2.5. Adicción**

Según la OMS (Organización Mundial de la Salud) [22], se entiende por adicción un estado de intoxicación crónica y periódica originada por el consumo repetido de una droga, natural o sintética, caracterizada por:

- Una compulsión a continuar consumiendo por cualquier medio.
- Una tendencia al aumento de la dosis.
- Una dependencia psíquica y generalmente física de los efectos.
- Consecuencias perjudiciales para el individuo y la sociedad.

### **2.5.1. Adicción al Teléfono Móvil**

La gran y cada vez más progresiva dependencia al móvil ha creado nuevas patologías, como la nomofobia (miedo irracional a estar sin el teléfono móvil). Esta dependencia genera una injustificable sensación de incomunicación en la



persona cuando no puede disponer de él, bien porque esté fuera de cobertura, porque lo haya dejado olvidado en casa, o porque se esté quedando sin batería o lo haya agotado. [23]

El promedio de consulta del móvil se tenga o no alguna notificación es de 150. Varios estudios realizados sobre este tema reflejan que cada vez se hace más complicada la vida sin el teléfono móvil, ya que actualmente se utiliza para todo y en algunos casos los usuarios basan su vida en él [23]. Un estudio realizado por Deloitte Consulting S.L.U. sobre el uso que hacen los españoles de los dispositivos móviles indica que el 29% de los jóvenes, de edad comprendida entre los 18 y los 24 años [24].

Los usos más frecuentes de los teléfonos móviles son: la cámara, las llamadas, el reloj (para consultar la hora o como despertador), chatear, buscar información a través de Internet y redes sociales. Este uso abusivo del móvil llega a tal extremo, que es constante, aunque se esté realizando otra actividad como ver la televisión, conducir o comer. La dependencia es tal, que la mayor parte de la población lo primero que hace nada más levantarse y lo último antes de acostarse es mirar su dispositivo móvil. [24].

## **2.5.2. Adicción a las Redes Sociales**

Las redes sociales han logrado cambiar la forma en la que las personas se relacionan y expresan con los demás. Por un lado, puede resultar beneficioso, ya que una interacción social es positiva (se conoce fácilmente gente nueva, resulta más fácil para retomar la relación con un conocido del pasado o de la infancia...), pero la preocupación se centra en los problemas psicológicos que estas plataformas pueden evidenciar y que podrían acentuarse con estas formas de interacción (un simple “Me gusta” que no te hace un amigo, un rechazo de una petición de amistad, ver una foto de amigos sin haber sido avisado a tal encuentro...) [25] [26].

Actualmente, el número de personas que acceden a diferentes plataformas de redes sociales asciende a más de 1.200 millones. Por tanto, se podría decir que el uso de las redes sociales se ha convertido casi en un estilo de vida, ya que todo cambio importante ocurrido en la vida del usuario es reflejado en una red social, bien sean fotografías, actualización de estados o publicación de vídeos. Sin importar la forma y el lugar, sólo la actualización de la información constante, y es precisamente este “importa” el que ha hecho que para la gran mayoría se convierta en una necesidad prioritaria y preocupante. Las redes sociales han logrado llegar a ser un medio de comunicación eficiente, logrando acercar a personas que están lejos. Sin embargo, también puede alejar a personas más cercanas al usuario por ser un inteligente distractor. Por ejemplo, caso muy habitual hoy en días es ver un grupo de amigos en un bar más interesado en su dispositivo móvil que en sus amigos [26].

## **2.6. Uso de Imágenes y Vídeos en las Redes Sociales**

Subir vídeos e imágenes desde los dispositivos móviles es la nueva tendencia en las redes sociales, ya que son el nexo de unión entre los usuarios y las redes sociales a las que están suscritos. Los usuarios comparten en las redes sociales sus viajes, eventos a los que asisten, opinión e incluso sentimientos, con el resto de usuarios.

Si a la rápida evolución de los teléfonos móviles, la mejora de las cámaras trasera y delantera se le suma la facilidad de conectar estos dispositivos a las redes sociales a través de aplicaciones móviles, se le da un gran incentivo a los para usar en cualquier momento y lugar. Esto, ha originado la nueva moda de tomar “selfies” y compartirlas en las redes sociales.

Una manera en la que un usuario se puede dar a conocer y opinar es compartiendo fotos, vídeos y opiniones en las redes sociales. Donde cada “me gusta”, foto o vídeo compartido son una forma de reconocimiento hacia este

usuario que las comparte.

Asimismo, los álbumes de fotos que almacenaban recuerdos familiares, de viajes y amigos, han pasado a un segundo plano y en vías de desaparecer. Esto no pasa con el uso de las redes sociales, ya que las fotos o vídeos generados con un dispositivo móvil tienen altas probabilidades de ser compartido directamente en las redes sociales y a su vez, estas publicaciones ser compartidas por otros usuarios.



### 3. ESTADO DEL ARTE

---

En este capítulo se describen las principales técnicas de análisis forense de imágenes digitales para la identificación de la fuente de adquisición de la imagen y los principales trabajos del estado del arte.

Estas técnicas tienen como objetivo identificar marca y modelo, de los dispositivos utilizados para la generación de imágenes digitales. Su éxito depende del supuesto de que todas las imágenes adquiridas por un mismo dispositivo presentan características intrínsecas del mismo. Las características utilizadas para la identificación de marca y modelo se derivan de las diferencias que existen entre las técnicas de procesamiento de las imágenes y las tecnologías de los componentes que se utilizan [27][28] en su proceso de captura. El mayor problema que tienen estas técnicas es que los diferentes modelos de las cámaras digitales usan componentes de un número reducido de fabricantes, y que los algoritmos que usan también son muy similares entre modelos de la misma marca.

Dentro de la identificación de la fuente existen dos grandes enfoques: escenarios cerrados o escenarios abiertos. Un escenario cerrado es aquel en el cual la identificación de la fuente de la imagen se realiza sobre un conjunto de cámaras concreto y conocidas a priori. Para este enfoque normalmente se utiliza un conjunto de imágenes de cada cámara para entrenar un clasificador y posteriormente se predice la fuente de adquisición de las imágenes objeto de investigación. La técnica más utilizada para la tarea de clasificación de imágenes digitales son las máquinas de soporte vectorial o SVM (del inglés *Support Vector Machine*).

Este trabajo se centra en la identificación de la fuente en escenarios abiertos, es decir, el analista forense no conoce a priori el conjunto de cámaras a las que

pertenece la imagen a identificar su fuente. En este tipo de clasificación, en la que no se tienen datos de cámaras a priori, el objetivo no es identificar la marca y modelo de la cámara, sino poder agrupar distintas imágenes en grupos disjuntos en los que todas sus imágenes pertenecen al mismo dispositivo.

### 3.1. Técnicas de Identificación de la Fuente

Estas técnicas se pueden clasificar dependiendo de la fase de generación de la imagen en la que se enfoque [27]: Utilización de la aberración de las lentes, interpolación de la matriz CFA, uso de las características de la imagen e imperfecciones del sensor. Además de los anteriores grupos existe otro grupo de técnicas forenses a destacar basadas en los metadatos de la imagen.

- **Técnicas Basadas en metadatos:** Las cámaras digitales cuentan con una poderosa fuente de información que son los metadatos embebidos en los archivos de las imágenes digitales. Los metadatos o “datos sobre datos” registran información relacionada con las condiciones de captura de la imagen, como fecha y hora de generación, presencia o ausencia de *flash*, distancia de los objetos, tiempo de exposición, apertura del obturador e información GPS entre otros. En otras palabras, información de interés que complementa el contenido principal de un documento digital. La especificación Exif es el contenedor de metadatos más común en las cámaras digitales [29]. La especificación Exif incluye cientos de etiquetas, entre las que se encuentran *marca* y *modelo*, aunque cabe destacar que la propia especificación no hace obligatoria su existencia en los archivos de las imágenes. Existen gran cantidad de trabajos enfocados sobre los diferentes tipos de metadatos, tanto para la búsqueda de información, como para la clasificación de imágenes [30][31]. Estas técnicas son las más sencillas, pero dependen en gran medida de los metadatos que los fabricantes deciden insertar cuando la imagen es generada y en la corrección de los mismos.

Asimismo, este método es el más vulnerable a modificaciones malintencionadas.

- **Técnicas Basadas en la Aberración de las Lentes:** Se basan en la idea de introducir aberraciones en la parte del sistema de lentes durante el proceso de generación de la imagen. Existen diferentes tipos de aberraciones: distorsión radial y distorsión cromática, esférica, coma, astigmatismo, curvatura de campo, etc. La distorsión radial es el tipo de aberración que más consecuencias tiene sobre la imagen, especialmente en las cámaras que usan lentes baratas de gran angular. La mayoría de cámaras digitales utilizan este tipo de lentes por cuestiones de coste [32].
- **Técnicas Basadas en la Interpolación de la Matriz CFA:** La elección de la matriz de colores CFA y la especificación de los algoritmos de interpolación cromática generan algunas de las diferencias más marcadas entre los diferentes modelos de cámaras [33][34][35][36]. Las cámaras comerciales tienen un solo sensor en lugar de varios sensores para cada componente del color. En esencia, la interpolación cromática introduce un tipo específico de correlación entre los valores de colores de los píxeles de la imagen. La forma específica de estas dependencias se puede extraer de las imágenes para diferenciar los algoritmos de interpolación cromática y así determinar marca y modelo de la cámara que generó una imagen.
- **Técnicas Basadas en las Características de las Imágenes:** Estas técnicas utilizan un conjunto de características extraídas del contenido de la imagen para hacer la identificación de la fuente. Estas características se dividen en tres grupos: características de color, métricas de calidad de la imagen IQM (del inglés *Image Quality Metrics*) y estadísticas del dominio wavelet [37][38][39][40].
- **Técnicas Basadas en las Imperfecciones del Sensor:** Estas técnicas se basan en el estudio de las huellas que los defectos del sensor dejan sobre las imágenes. Estas técnicas se dividen en dos ramas: defectos de píxel y patrón de ruido del sensor SPN (del inglés, *Sensor Pattern Noise*). En la primera se

estudian los defectos de píxel, los píxeles calientes, los píxeles muertos, los defectos de fila o columna, y los defectos de grupo. En la segunda se construye un patrón del ruido promediando los múltiples residuos de ruido obtenidos mediante algún filtro de eliminación de ruido. La presencia del patrón se determina utilizando algún método de clasificación como correlación o máquinas SVM [41][42][43].

Este trabajo utiliza las técnicas basadas en las imperfecciones del sensor, concretamente aquellas basadas en el patrón de ruido del sensor (SPN) el cual es originado por las imperfecciones en el proceso de fabricación de los semiconductores o las producidas por la utilización de la cámara en el día a día.

### **3.2. Técnicas de Clustering**

El objetivo del análisis de *clústeres* o *clustering* es agrupar una colección de objetos en clases representativas llamadas *clústeres*, sin información a priori, de forma que los objetos pertenecientes a cada *clúster* guarden una mayor similitud con respecto de objetos en otros *clústeres*. La agrupación de imágenes puede llevarse a cabo mediante técnicas de aprendizaje supervisadas o sin supervisión. En el primer caso es indispensable conocer información del dispositivo a priori, es decir se identifica claramente con la clasificación en escenarios cerrados en donde se requiere una fase de entrenamiento con las características extraídas de las imágenes y una segunda fase de clasificación conforme al resultado anterior. Sin embargo, en un caso real puede ser difícil contar con la cámara en cuestión o con un subconjunto de fotografías tomadas por la misma para llevar a cabo un entrenamiento, de ahí la necesidad de técnicas de aprendizaje sin supervisión, que se corresponden directamente con los escenarios abiertos. El *clustering* tradicional se caracteriza por ser una técnica de aprendizaje sin supervisión.



Para poder determinar la similitud entre objetos pertenecientes a un mismo *clúster* existen medidas de distancia como pueden ser: distancia euclidiana, distancia Manhattan y distancia Chebychev, entre otras. Alternativamente, es posible usar funciones de similitud que comparan dos vectores en forma simétrica. Estas funciones alcanzan sus valores más altos cuando son más similares. La medida más usada en la identificación de fuente de imágenes es la correlación normalizada [44][45][46].

De acuerdo a la clasificación de algoritmos de *clustering* propuesta en [47] encontramos los métodos jerárquicos cuyo propósito es lograr una estructura denominada dendograma que representa la agrupación de los objetos de acuerdo a sus niveles de similitud. Esta agrupación puede realizarse de distintas formas: aglomerativa o divisiva.

- **Agrupación aglomerativa:** considera inicialmente a cada objeto como una clase independiente hasta, de forma iterativa, lograr agrupar todos los objetos en una clase única.
- **Agrupación divisiva:** se basa en la idea de partir de una sola clase hasta lograr separar todos los objetos en clases individuales.

También existen los algoritmos de particionamiento en donde iniciando de una partición, el algoritmo se encarga de mover objetos de un *clúster* a otro hasta minimizar cierto criterio de error. Dentro de esta categoría el método más famoso es el k-means, sin embargo, la mayoría de estos métodos requieren conocer de antemano el número de *clústeres*, por lo cual no son muy utilizados en temas de análisis forense de imágenes. Existen trabajos previos sobre agrupación de imágenes por métodos sin supervisión, todos ellos consideran al SPN como el criterio más fiable para representar la huella digital de un dispositivo, es de ahí que utilizan concretamente el PRNU (del inglés *Photo Response Non-Uniformity*) como huella y la correlación normalizada como medida de similitud para lograr el agrupamiento de imágenes por dispositivo.



## 4. THEIA: HERRAMIENTA PARA EL ANÁLISIS FORENSE DE IMÁGENES Y VÍDEOS DIGITALES

En este capítulo se presenta Theia, una herramienta para el análisis forense de imágenes y vídeos digitales que facilita la extracción y el tratamiento de metadatos Exif en imágenes JPEG y átomos en vídeos MP4. A grandes rasgos la herramienta se divide en dos grandes partes: Tratamiento individual y tratamiento masivo de imágenes y vídeos.

### 4.1. Tratamiento a nivel individual

Permite obtener la información Exif detallada de una imagen individual, situar la imagen en Google Maps y Google Earth (si posee información de geoposicionamiento). La estructura general se puede observar en la Figura 4.1.

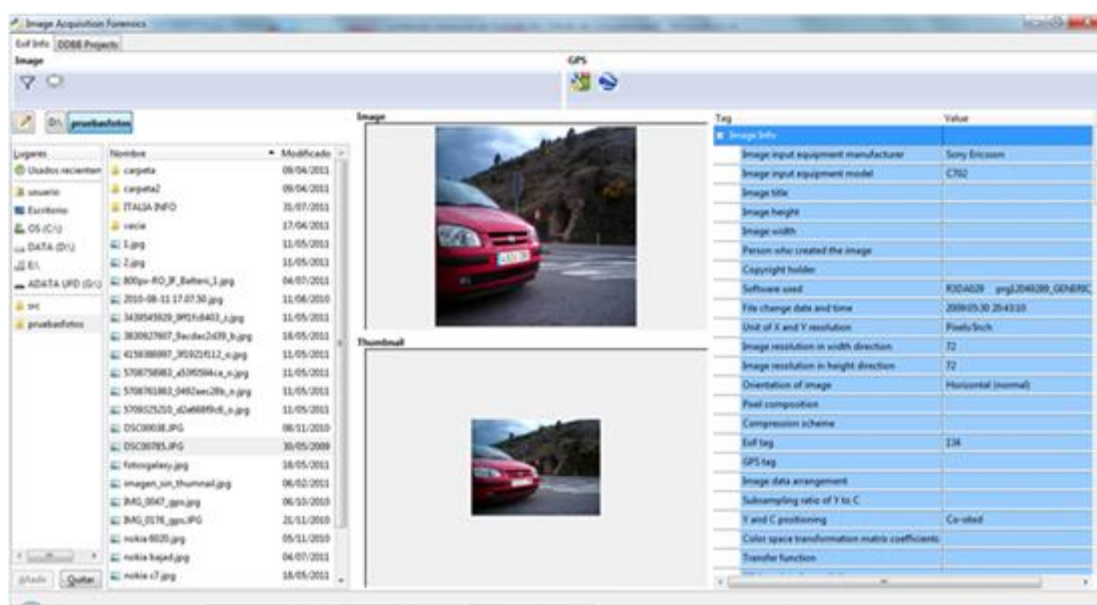


Figura 4.1. Apariencia general de la pestaña *Exif Info*

Como estructura general se puede apreciar a la izquierda de la imagen un navegador de archivos, en el centro la imagen del archivo seleccionado y su correspondiente thumbnail (es el incluido en el propia archivo de la imagen no ninguna generación propia del programa) y a la derecha las etiquetas Exif con

su correspondiente información. De esta estructura cabe destacar en la interfaz gráfica que es totalmente configurable a nivel de tamaños, es decir todos los separadores entre las distintas zonas se pueden mover.

La información Exif se ha organizado en 6 grupos: *Image*, *Exif*, *GPS*, *Interoperability*, *Thumbnail* y *Maker Note*.

- ***Image Info***: En este bloque se almacenan las etiquetas con información relativa a la propia imagen y que no tienen relación directa con el entorno y el momento de la captura. Por ejemplo la marca y modelo de la cámara, el tamaño de la imagen, la unidad utilizada en la resolución X e Y, etc.
- ***Exif Info***: En este bloque se guardan las etiquetas con información relativa al momento o al entorno de la toma de la imagen. Dentro de este bloque se encuentra por ejemplo la información referente al flash, hora de toma y generación de la imagen, configuración de la lente, etc.
- ***GPS Info***: En este bloque está toda la información relativa al geoposicionamiento. Por ejemplo información de latitud, longitud, altitud, el estado del receptor GPS, etc.
- ***InterOperability Info***: En este bloque se incluyen las etiquetas relativas a la información de las reglas de interoperabilidad, como pueden ser *Exif R98*, *DCF thumbnail file* o *DCF Option file*.
- ***Thumbnail Info***: En este bloque se encuentran todas las etiquetas relativas a la información de *thumbnail*. Por ejemplo su tamaño en vertical y horizontal y el esquema de compresión utilizado.
- ***Maker Note Info***: Es una etiqueta individual que almacena la información que cada fabricante puede insertar de forma opcional y que no ha sido recogida en ninguna etiqueta Exif.

El formato de esta información es libre y no tiene una estructura prefijada, cada fabricante utiliza la suya propia que incluso puede ser diferente para distintos modelos de la misma marca. Por tanto se muestra como una secuencia de bytes (en hexadecimal). Si se conoce la estructura estos bytes pueden ser decodificados de forma manual.

Asimismo, se pueden extraer los datos de geoposicionamiento que estén incluidos en las imágenes. Si la imagen no tiene la suficiente información para poder ser mostrada en alguna de las opciones al pulsar la opción de geoposicionamiento se mostrará un mensaje indicándolo (*Not enough GPS information*). El geoposicionamiento se puede realizar desde dos opciones: posicionamiento en Google Maps y en Google Earth. En la primera se abrirá el navegador web por defecto del sistema operativo y se mostrará la ubicación inserta en los metadatos de la imagen en un mapa de Google Maps (es necesario conexión a internet). La Figura 4.2 muestra un ejemplo de geoposicionamiento en una fotografía en Google Maps.

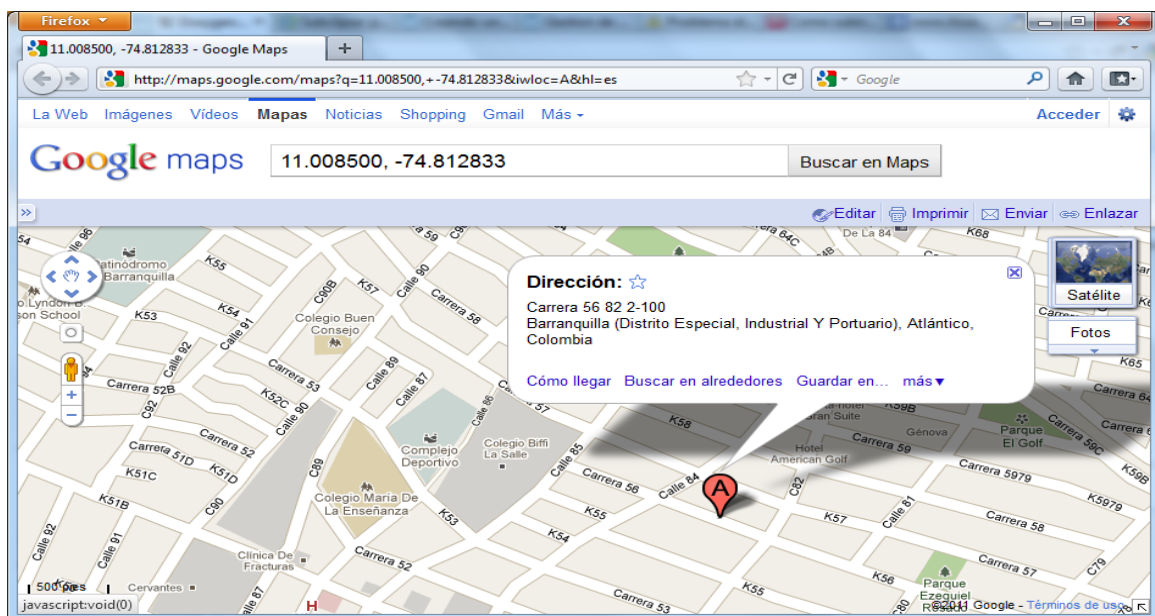
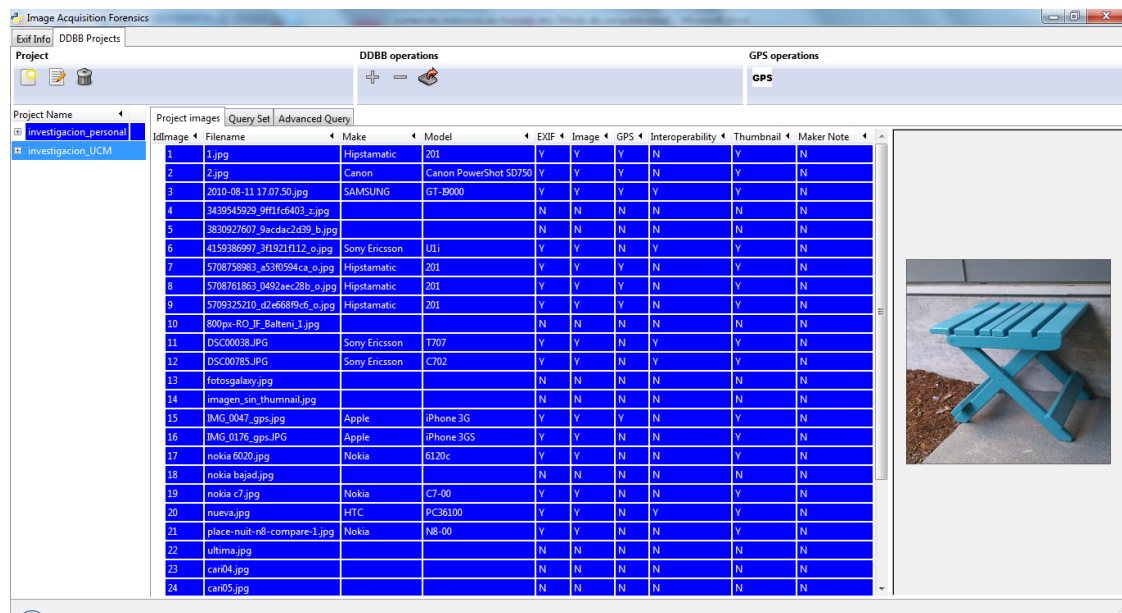


Figura 4.2. Geoposicionamiento en Google Maps

En la segunda opción se abrirá un menú para poder almacenar un archivo de extensión “kml”. Este archivo podrá ser posteriormente abierto si está instalada la aplicación Google Earth, en la cual se mostrará igualmente la posición geográfica almacenada en los metadatos de la imagen (es necesario conexión a internet).

## 4.2. Tratamiento a nivel de grupo

Permite hacer análisis de imágenes en grupo. Cada grupo es totalmente independiente entre sí. Su apariencia gráfica general puede verse en la Figura 4.3.



Project Name	Project images	Query Set	Advanced Query
investigacion_personal			
investigacion_UCM			

IdImage	Filename	Make	Model	EXIF	Image	GPS	Interoperability	Thumbnail	Maker Note
1	1.jpg	Hipstamatic	201	Y	Y	Y	N	Y	N
2	2.jpg	Canon	Canon PowerShot SD750	Y	Y	Y	N	Y	N
3	2010-08-11 17:07:50.jpg	SAMSUNG	GT-B000	Y	Y	Y	Y	Y	N
4	3439545929_9m1fc403_z.jpg			N	N	N	N	N	N
5	3830927607_9acdac2d39_b.jpg			N	N	N	N	N	N
6	4159386997_3f1921f112_o.jpg	Sony Ericsson	UIi	Y	Y	N	Y	Y	N
7	5708758983_a53f0594ca_o.jpg	Hipstamatic	201	Y	Y	Y	N	Y	N
8	5708761863_0492ae28b_o.jpg	Hipstamatic	201	Y	Y	Y	N	Y	N
9	5709325210_d2e668f8d_o.jpg	Hipstamatic	201	Y	Y	Y	N	Y	N
10	800px-ROJF_BaRen1.jpg			N	N	N	N	N	N
11	D5C00038.JPG	Sony Ericsson	T707	Y	Y	N	Y	Y	N
12	D5C00785.JPG	Sony Ericsson	C702	Y	Y	N	Y	Y	N
13	fotosgalaxy.jpg			N	N	N	N	N	N
14	imagen_sin_thumbnail.jpg			N	N	N	N	N	N
15	IMG_0647_gps.jpg	Apple	iPhone 3G	Y	Y	Y	N	Y	N
16	IMG_0176_gps.JPG	Apple	iPhone 3GS	Y	Y	N	N	Y	N
17	nokia 6020.jpg	Nokia	6120c	Y	Y	N	N	Y	N
18	nokia bajad.jpg			N	N	N	N	N	N
19	nokia c7.jpg	Nokia	C7-00	Y	Y	N	N	Y	N
20	nueva.jpg	HTC	PC36100	Y	Y	N	Y	Y	N
21	place-nuit-rb-compare-1.jpg	Nokia	N8-00	Y	Y	N	N	Y	N
22	ultima.jpg			N	N	N	N	N	N
23	can04.jpg			N	N	N	N	N	N
24	can05.jpg			N	N	N	N	N	N

Figura 4.3. Apariencia general de la pestaña *DDBB Projects*

Lo primero a destacar en esta funcionalidad es que las imágenes se tratan en grupos llamados proyectos. Estos grupos pueden ser de una o más imágenes. Cada proyecto es totalmente independiente entre sí. Se busca acercar la realidad del día a día del analista forense a la herramienta, es decir, el analista tendrá diversos casos de análisis disjuntos los cuales podrá tratar en proyectos distintos.

En la parte central de la pestaña *DDBB Projects* y dentro de ésta en la pestaña *Project Images* se muestran una lista de las imágenes del proyecto seleccionado en la lista de proyectos.

Para cada imagen se muestra su identificador interno de la base de datos (para permitir el caso de archivos con el mismo nombre), el nombre del archivo, la marca y el modelo de dispositivo que la creó (si existe). Además se presenta la información de si posee metadatos en los distintos grupos Exif que analiza la herramienta. Asimismo se visualiza el contenido de cada una de las imágenes a la derecha según se van seleccionando. Un ejemplo de captura de esta funcionalidad se muestra en la Figura 4.4.

Project images									
Query Set Advanced Query									
IdImage	Filename	Make	Model	EXIF	Image	GPS	Interoperability	Thumbnail	Maker Note
32	DSC00398.JPG	Sony Ericsson	W580i	Y	Y	N	Y	Y	N
33	DSC00403.JPG	Sony Ericsson	W580i	Y	Y	N	Y	Y	N
34	DSC00404.JPG	Sony Ericsson	W580i	Y	Y	N	Y	Y	N
35	DSC00414.JPG	Sony Ericsson	W580i	Y	Y	N	Y	Y	N
36	DSC00415.JPG	Sony Ericsson	W580i	Y	Y	N	Y	Y	N
37	DSC00416.JPG	Sony Ericsson	W580i	Y	Y	N	Y	Y	N
38	DSC00398.JPG	Sony Ericsson	W580i	Y	Y	N	Y	Y	N

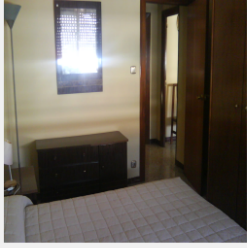
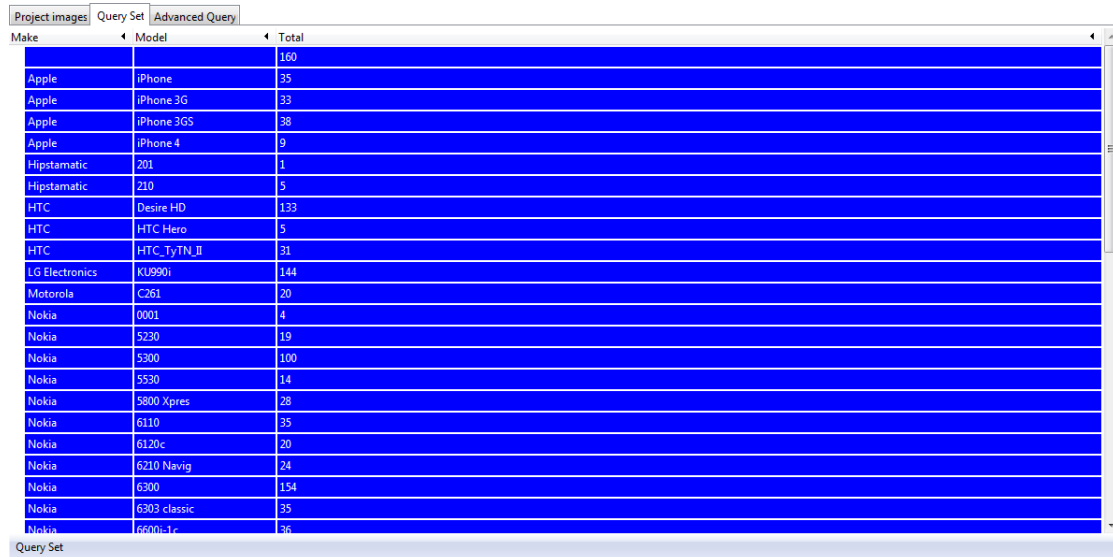


Figura 4.4. Visualización de las imágenes de un proyecto

Los diferentes análisis que se pueden realizar sobre cada proyecto son los siguientes: administración de imágenes (añadir y eliminar imágenes), consultas preestablecidas, consultas avanzadas y geoposicionamiento de las imágenes.

- *Consultas preestablecidas:* Permite crear consultas agregando etiquetas Exif (y otras adicionales que añade la aplicación que ayudan al análisis forense) sobre las imágenes del grupo seleccionado. La consulta agrupa las imágenes por los criterios seleccionados y muestra el número de imágenes que hay en

cada uno de los grupos formados, como puede verse en un ejemplo en la Figura 4.5. En las consultas permiten escoger 5 campos de agregación como máximo (por defecto se realiza sobre *Make* y *Model*, aunque se pueden elegir cualesquiera). Para escoger los distintos campos hay que pulsar sobre el botón *Query Set*



Make	Model	Total
		160
Apple	iPhone	35
Apple	iPhone 3G	33
Apple	iPhone 3GS	38
Apple	iPhone 4	9
Hipstamatic	201	1
Hipstamatic	210	5
HTC	Desire HD	133
HTC	HTC Hero	5
HTC	HTC_TyTN_II	31
LG Electronics	KU990i	144
Motorola	C261	20
Nokia	0001	4
Nokia	5230	19
Nokia	5300	100
Nokia	5530	14
Nokia	5800 Xpres	28
Nokia	6110	35
Nokia	6120c	20
Nokia	6210 Navig	24
Nokia	6300	154
Nokia	6303 classic	35
Nokia	6600i-1c	36

Figura 4.5. Query Set

- *Consultas avanzadas*: Permite la creación de consultas sobre imágenes de un grupo configurando los datos Exif a mostrar y los filtros a aplicar. Es decir, muestra la información de las imágenes de los campos seleccionados que coincidan con uno de los valores de cada uno de los filtros configurados. Asimismo, se permite el almacenamiento permanente de consultas. Una visión general se muestra en la Figura 4.6.

En *Advanced Query* hay que distinguir dos grandes bloques: la configuración de la consulta y su almacenamiento. Con respecto a la configuración de la consulta avanzada hay que tener en cuenta la configuración de las columnas de los resultados y la configuración de los filtros. En esta consulta se muestran los valores de los campos seleccionados por la configuración de las columnas de los resultados que cumplen las restricciones indicadas en la configuración de los filtros.



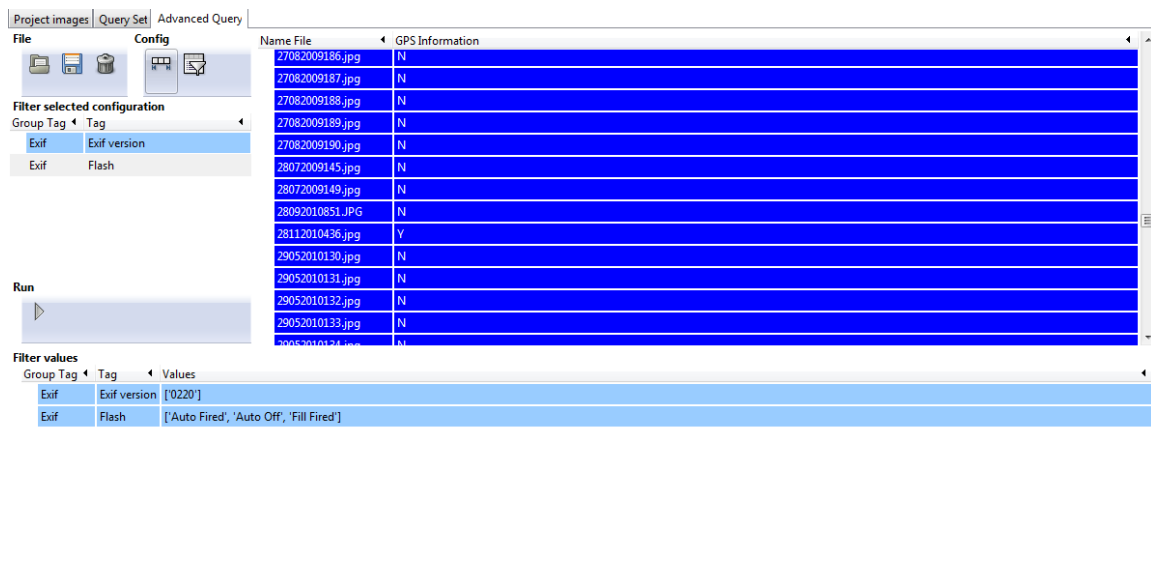


Figura 4.6. *Advanced Query*

- *Geoposicionamiento*: Análogamente al tratamiento de imágenes a nivel individual, existe una funcionalidad que permite el tratamiento de la información de geoposicionamiento para un grupo de imágenes. Esta opción permite la selección de algunas o de todas las imágenes de un grupo con información de geoposicionamiento para la creación de un mapa en Google Maps que sitúe a las mismas. En el mapa se agrupan las imágenes por zona y, a medida que se aumenta el *zoom*, se van detallando las coordenadas. La Figura 4.7 muestra un ejemplo del mapa generado y el proceso de aumento del *zoom* en una zona concreta (desde la Figura 4.7 (a) hasta la Figura 4.7 (d)).



(a)



(b)



(c)



(d)

Figura 4.7. Geoposicionamiento de un grupo de imágenes en Google Maps

## 5. CONTRIBUCIÓN

---

En este trabajo se ha diseñado un algoritmo de identificación de la fuente de imágenes de dispositivos móviles de imágenes descargadas de las redes sociales Facebook y Flickr. Asimismo, se ha implementado un par de procesos para descarga automática de imágenes desde las redes sociales Facebook y Flickr. Por razones de confidencialidad del proyecto se ha omitido información del trabajo desarrollado para no infringir la normativa correspondiente.

### 5.1. Consideraciones Generales

Este trabajo se basa en el supuesto de que las cuentas que un usuario tiene en una determinada red social contienen imágenes que han sido tomadas con el dispositivo móvil del titular de la cuenta y que ha subido a la red social directamente desde su propio dispositivo móvil.

Cuando una foto es subida a una red social se expone a ser modificada y a la manipulación de sus metadatos como mecanismos de optimizar los recursos de almacenamiento de la red social. Modificaciones como redimensión automática, supresión de los metadatos, compresión JPEG adicional, etc. para reducir el tamaño del fichero. Este es el caso de Facebook, Flickr y Twitter, entre otras. Esto no ocurre con Google+ que permite subir las imágenes en su versión original. Asimismo, a la hora de descargar una imagen de una red social, algunas de estas redes sociales permiten elegir el tamaño de descarga de las imágenes, incluyendo la imagen tal como fue subida.

Facebook permite elegir el formato de subida, permitiendo subir la foto en tamaño original, pero al procesarla la redimensiona automáticamente. A la hora de descargar la imagen no permite modificar la resolución del fichero. Flickr, en cambio, no muestra ninguna opción de procesamiento al momento de subir la

imagen. Sin embargo, almacena diferentes formatos de la imagen, dejando entre ellos el formato subido originalmente. En el proceso de descarga de la imagen, permite elegir la resolución de la imagen entre las que se encuentra la resolución original. Todas estas modificaciones realizadas por las redes sociales en las fotos durante los procesos de carga y descarga de las mismas, provoca una mayor dificultad a la hora del tratamiento de estas imágenes como pruebas científicas, ya que se pueden perder información importante para la identificación de la fuente de las imágenes.

## **5.2. Funcionamiento**

El algoritmo de agrupación sin supervisión implementado en este trabajo es una combinación entre un clustering jerárquico y un clustering plano. Es decir, a pesar de formar una estructura de dendograma con cada iteración del algoritmo, al final los clústeres son tomados como entidades sin relación alguna ya que cada uno de ellos debe corresponder a un dispositivo específico. Previo a realizar el clustering, es necesario obtener los patrones de ruido del sensor del conjunto de imágenes.

En la Figura 5.1 se presenta el esquema general de funcionamiento del algoritmo.

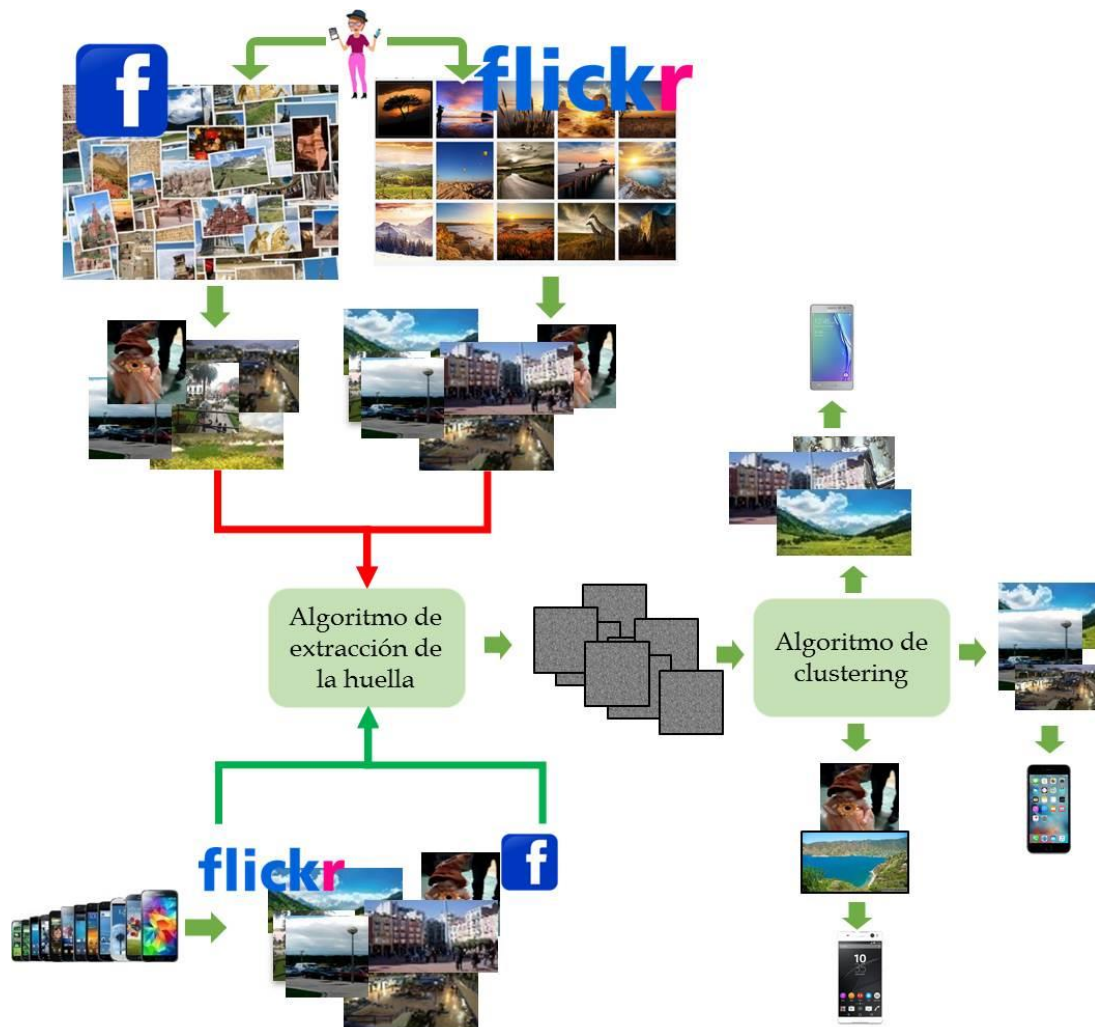


Figura 5.1 Esquema general de funcionamiento del algoritmo

Como se mencionó anteriormente, el objetivo del clustering es agrupar objetos en un entorno no supervisado (escenario cerrado), sin embargo es factible llevar a cabo una etapa de entrenamiento y una de clasificación. Esto permite reducir la complejidad computacional y por ende reducir el tiempo de ejecución del algoritmo. Para conseguir este objetivo, se cuenta con dos conjuntos de imágenes: uno para el entrenamiento y otro para la clasificación.

Las imágenes utilizadas en el entrenamiento son imágenes que han sido tomadas con varios dispositivos móviles entre los que se encuentra el dispositivo investigado. Posteriormente subidas y descargadas de las mismas redes sociales que se van a analizar. Seguidamente se extrae el ruido SPN de

cada imagen. Con este conjunto se obtiene el número de clústeres (K) que representa a cada dispositivo. Luego se calcula un centroide por cada clúster promediando todos los patrones de ruido contenidos en éste. A continuación se obtiene el valor de correlación entre cada SPN del conjunto y cada centroide.

El conjunto de fotos utilizadas en la clasificación se obtiene descargando las fotos del usuario investigado desde las redes sociales Facebook y Flickr. Seguidamente, como en el caso de las imágenes entrenadas se extrae el ruido SPN de cada imagen. A continuación, Cada imagen es clasificada en el clúster donde haya existido un valor mayor de correlación. La correlación se calcula utilizando la formula (1).

$$correlación(R_1, R_2) = \frac{\sum_i R_1'(i) \cdot R_2'(i)}{\sqrt{\sum_i R_1'^2(i) \cdot R_2'^2(i)}} \quad (1)$$

Para la correlación, un resultado cercano a 1 (correlación directa) representa una mayor similitud y para una baja similitud cercana a -1 (correlación inversa), el valor de 0 indica que no existe una correlación (asociación aleatoria).

Para la implementación de los algoritmos de descarga automática de fotos, se han elegido dos de las redes sociales más utilizadas por los usuarios para compartir imágenes y vídeos. Estas redes sociales son: Facebook y Flickr.

Para poder descargar las fotos de un álbum del perfil de un usuario perteneciente a una red social como Facebook o Flickr es necesario:

1. Tener un usuario registrado en la red social.
2. Obtener unas credenciales o “token” con permisos especiales otorgados por los desarrolladores de la red social. En Facebook se solicita unas credenciales con permisos especiales y en Flickr una clave pública y una privada para utilizar las APIs de cada red social.
3. Obtener una autorización expresa del usuario investigado o de una

entidad judicial para tener acceso a los datos contenidos en la red social, en este caso las fotos.

El primer paso del proceso es iniciar una sesión en Facebook con un usuario que posteriormente realizará la petición del token. Hay que tener en cuenta que esta sesión debe mantenerse activa durante todo el proceso de descarga. El objetivo del token es otorgarle permisos de acceso como desarrollador a la API de Facebook. Posteriormente, se obtiene el listado de la información que se quiere descargar, en este caso las se desean descargar las fotos del álbum, por lo que los parámetros son: el identificador del álbum y “photos” que es la información deseada del álbum.

Análogamente se sigue el mismo proceso en Flickr, se obtiene el identificador del usuario, álbum y la resolución de las fotos que se desean descargar.

### 5.3. Herramientas de Desarrollo

Para el desarrollo de este trabajo se evaluaron y utilizaron las siguientes herramientas:

- **SeleniumHQ** [49]: Herramienta de automatización web que facilita la ejecución de pruebas sobre aplicaciones web. En un primer momento se pensó en utilizarla para la descarga de Fotos de las redes sociales. Sin embargo, encontró la limitación de no soportar cambios en la interfaz web analizada provocando que el código del algoritmo no sea estable.
- **Python**: Se ha utilizado el lenguaje de programación Python en su versión 2.7.6 ya que tiene una licencia de código abierto (*Python Software Foundation License*). Esta licencia es compatible con la Licencia Pública General de GNU a partir de la versión 2.1.1

- **API Graph de Facebook:** Es una librería que permite acceder a la red social para introducir y extraer datos. Es una API basada en HTML a bajo nivel para consultar datos, publicar nuevas historias, subir fotos y muchas otras tareas. Fue utilizada para la implementación del algoritmo de descarga de fotos de Facebook ha sido necesaria la librería de Facebook
- **Flickr API:** Es propiedad de Flickr y su compañía madre Yahoo! Inc. Está compuesta por un grupo de métodos que permiten acceder a los datos de la red social como desarrollador. Tiene una licencia de uso gratis y una de uso comercial condicionadas a unas reglas de uso legal del contenido de los usuarios

## 5.4. Evaluación de los Algoritmos

Los experimentos realizados tienen un doble propósito: Realizar un estudio sobre los procesos de filtrado y compresión aplicados por cada red social a las imágenes subidas a cada red social y ejecutar el algoritmo de identificación de la fuente para comprobar la efectividad de los algoritmos propuestos con imágenes procedentes de diferentes redes sociales y que fueron tomadas con el mismo dispositivo móvil.

### 5.4.1. Configuración de los Experimentos

En el desarrollo de este trabajo se han utilizado 3 de las redes sociales más importantes como objeto de estudio: Facebook, Flickr y Google +. Estas redes sociales tienen dos características que las hicieron ideales para su estudio: Permiten subir y descargar imágenes y vídeos, y dos de ellas tienen librerías en Python que pueden ser utilizadas para desarrollar scripts para acceder a los datos.

Los experimentos fueron realizados con un conjunto de 3300 fotos de 33 modelos diferentes de cámaras de dispositivos móviles de 12 fabricantes. En la



Tabla 4.1 se presentan los fabricantes y modelos utilizados para los experimentos y la configuración de sus cámaras.

Marca	Modelo	Resolución	Condiciones de Captura
Apple	Iphone 4S	3264x2448	Tipo de escena: Cualquiera Orientación: Cualquiera Flash: Indiferente Luz: Natural Balanceo de blancos: Automático Zoom Digital: 0 Tiempo de exposición: 0 Velocidad ISO: Automática
	Iphone 5		
	Iphone 5S		
	Iphone 6		
BlackBerry	8520	1600x1200	
BQ	BQ Aquaris E 4,5	2448x3264	
	BQ Aquaris E5	2448x4352	
Huawei	Huawei Y635-L01	1920x2560	
	U8815	1944x2592	
LG	E400	1536x2048	
	Nexus 5	2448x3264	
	P760	1944x2592	
Motorola	Moto G 1	1944x2592	
	Moto G 2	2448x3264	
	NEXUS 6	3120x4160	
Nokia	800-lumia	3264x2448	
One Plus	OnePlusOne A0001	4160x3120	
Samsung	Galaxy Nexus	2592x1994	
	Galaxy S3	3264x2448	
	Galaxy S4 mini (GT-19 195)	3264x1836	
	Galaxy S5	5312x2988	
	Galaxy S6	5312x2988	
	GT-I9001	640x480	
	GT-I9100	3264x2448	
	GT-S5830	2560x1920	
	GT-S5830M		
	S3 NEO GT-I93011	3264x1836	
Sony Ericsson	C2105	3104x1746	
	ST25A	2560x1440	
	ST25I		
	XPERIA M2 D2303	3104x1746	
Xiaomi	MI3	4208x2368	
Zopo	ZP980	4096x3072	

Tabla 4.1. Dispositivos móviles utilizados en los experimentos

De cada modelo de dispositivo móvil se tomaron 100 fotos con el GPS activado. Las fotos de cada modelo fue subido a las 3 redes sociales con las resoluciones permitidas por cada una. Posteriormente, fueron descargadas las fotos con las resoluciones permitidas por cada red social. La Tabla 4.2 muestra el número de fotos resultante al finalizar la descarga de cada red social.

Red Social	Resolución Descargada	Fotos descargadas
Facebook	Normal	6600
	Alta	
Flickr	150x150	16500
	180x240	
	480x640	
	1536x2048	
	1920x2560	
Google +	Igual a la subida	3300
<b>Total</b>		<b>26400</b>

Tabla 4.2. Resoluciones de descarga disponibles en las redes sociales

En Facebook, el proceso de subida de fotos tiene dos opciones de resolución para subir las fotos: alta o normal. La resolución que utiliza realmente depende del tamaño original de la foto, sin sobrepasar 2048 píxeles en el tamaño normal y 1024 píxeles en el tamaño normal. Flickr, por el contrario, almacena las fotos en su resolución real y en el proceso de descarga permite elegir la resolución entre las presentadas en la Tabla 4.2. Siendo la resolución 1920x2560 considerada como original, a pesar de no serlo realmente. Por último, en Google+ las fotos son almacenadas y descargadas posteriormente en su resolución original, y no realiza modificaciones de las mismas. Una vez descargadas las imágenes de las tres redes sociales, éstas fueron recortadas a 1024x1024 píxeles para realizar los experimentos.

### 5.4.2. Experimentos

Se realizaron dos experimentos con el objetivo de evaluar el comportamiento del algoritmo de identificación bajo los siguientes escenarios: 1) Dispositivos de varios fabricantes y 2) Dispositivos del mismo fabricante. Para medir el grado de certeza en los resultados de los experimentos se utilizó la tasa de aciertos o TPR (del inglés *True Positive Rate*). La tasa de acierto promedio para los siguientes experimentos se calcula computando, para cada clúster o dispositivo móvil, el número de fotografías que han sido bien clasificadas (tasa de acierto de cada dispositivo móvil) y promediando la tasa de acierto de todos los dispositivos móviles.

En el primer experimento se utilizó un conjunto de fotografías de 11 modelos de 6 fabricantes: 90 fotografías diferentes descargadas de las 3 redes sociales (30 de cada una) con diferentes resoluciones, mayores a 1024 píxeles para la fase de entrenamiento y otras 90 fotografías para la fase de clasificación. Ninguna de las imágenes utilizadas en la fase de entrenamiento se encuentra en el conjunto de imágenes de la fase de clasificación y viceversa. La Tabla 4.3 presenta la matriz de confusión de este experimento. Se puede observar que la tasa de acierto es del 82,8% siendo el Iphone 6 el dispositivo que tenía el número de falsos positivos más alto (4% en promedio).

El segundo experimento se realizó sobre un conjunto de fotos de 5 modelos del mismo fabricante. El conjunto de fotos utilizado es similar a las fotos utilizadas en el experimento 1. Las cámaras de los dispositivos móviles de un mismo fabricante deberían ser muy similares para gran parte de sus productos y por tanto el ruido del sensor extraído entre modelos distintos debiera asemejarse. Sin embargo, la Tabla 4.4 muestra que la tasa de acierto ha aumentado a 88,8%. Por tanto, se concluye que para este experimento el valor de correlación entre SPN varía lo suficiente entre modelos para identificar a cada uno por separado, incluso cuando los modelos son muy parecidos.

Teléfono Móvil		Apple		Huawei	LG	Motorola		Samsung			Sony Ericsson	
		Iphone 5S	Iphone 6	U8815	Nexus 5	Moto G1	Moto G2	Galaxy S4	Galaxy S6	GTI 9100	C2105	ST25I
Apple	Iphone 5S	83%	11%	3%	0%	0%	1%	1%	1%	0%	0%	0%
	Iphone 6	5%	89%	3%	0%	0%	1%	0%	0%	1%	0%	0%
Huawei	u8815	3%	3%	87%	1%	0%	1%	0%	1%	2%	1%	1%
LG	Nexus 5	0%	0%	0%	88%	0%	0%	1%	2%	6%	1%	2%
Motorola	Moto G 1	1%	6%	4%	2%	73%	5%	0%	0%	0%	4%	5%
	Moto G 2	0%	8%	4%	12%	3%	72%	0%	0%	0%	0%	0%
Samsung	Galaxy S4	1%	1%	0%	3%	2%	1%	90%	2%	0%	0%	0%
	Galaxy S6	0%	3%	2%	1%	1%	0%	0%	93%	0%	0%	0%
	GT-I9100	2%	0%	4%	2%	0%	5%	0%	0%	86%	0%	1%
Sony Ericsson	C2105	2%	3%	0%	0%	3%	1%	3%	1%	1%	85%	0%
	ST25I	5%	6%	4%	1%	1%	2%	3%	1%	2%	0%	75%

Tabla 4.3. Matriz de Confusión del experimento 1 con una tasa de acierto del 82.8%

Teléfono Móvil		Samsung					
		Galaxy S3	Galaxy S5	Galaxy S6	GT-I9100	GT-S5830	GT-S5830M
Samsung	Galaxy S3	83%	11%	5%	0%	0%	1%
	Galaxy S5	6%	92%	0%	0%	2%	0%
	Galaxy S6	1%	2%	95%	1%	0%	1%
	GT-I9100	0%	4%	0%	88%	5%	3%
	GT-S5830	1%	6%	2%	1%	85%	5%
	GT-S5830M	0%	0%	3%	1%	6%	90%

Tabla 4.4. Matriz de Confusión del experimento 2 con una tasa de acierto del 88.8%

## 6. CONCLUSIONES Y TRABAJO FUTURO

---

### 6.1. Conclusiones

La mayoría de las redes sociales ofrecen a los usuarios la posibilidad de subir fotos para enriquecer su perfil. Algunas de estas redes sociales se dedican a compartir fotos y Vídeos.

Encontrar cuentas de redes sociales que pertenezcan a una determinada persona puede ser muy valiosa. La tarea es no trivial, incluso si se conoce detalles de la identidad. Sin embargo, desde la perspectiva de las investigaciones forenses digitales, identificar a la persona que generó una imagen es importante en varios escenarios. Por ejemplo, el caso de un teléfono inteligente robado, que después del robo, el ladrón lo utiliza para tomar fotografías y compartirlas en sus redes sociales. Las fotografías publicadas por la víctima en sus redes sociales pueden ser utilizadas como muestras para correlacionar las imágenes e identificar al ladrón.

En este trabajo se ha propuesto un algoritmo basado en el patrón PRNU, combinado con técnicas de agrupamiento, para identificar al propietario de un dispositivo móvil. La identificación de la fuente es realizada analizando las fotografías que un usuario de redes sociales sube desde el propio dispositivo.

La evaluación de la propuesta presentada se realizó con fotos subidas a Facebook, Flickr y Google +, 3 de las redes sociales más utilizadas actualmente. Para ello, se desarrollaron dos algoritmos para descargar de forma automática fotografías de estas redes sociales. Los experimentos se realizaron con un conjunto de 3300 fotografías de 33 modelos diferentes de cámaras de dispositivos móviles de 12 fabricantes, que fueron almacenados en las redes sociales y posteriormente descargados con las diferentes resoluciones disponibles en cada red social.

En los resultados de los experimentos se obtuvo una tasa de aciertos del 82,8% en la clasificación de 11 modelos de dispositivos móviles de 5 diferentes fabricantes y un 88,8% de tasa de acierto en la clasificación de fotos de fotografías de 5 modelos del fabricante Samsung. Las cámaras de los dispositivos móviles de un mismo fabricante deberían ser muy similares para gran parte de sus productos y por tanto el ruido del sensor extraído entre modelos distintos debiera asemejarse. Sin embargo, se observa que el valor de correlación entre el patrón PRNU varía lo suficiente entre modelos para identificar a cada uno por separado, incluso cuando los modelos son muy parecidos.

## **6.2. Trabajo Futuro**

Como trabajo futuro pueden señalarse las siguientes líneas de investigación:

- Ampliar la experimentación para evaluar el uso del patrón PRNU en resoluciones inferiores a 1024.
- Evaluar el uso de otros métodos de identificación de la fuente como los basados en las características de la imagen para identificar el propietario de una fotografía
- Ampliar los algoritmos de identificación de la fuente a vídeos almacenados en redes sociales.

## **RESUMEN EN INGLÉS**





## 7. INTRODUCTION

---

### 7.1. Motivation

Since the release of the first mobile phone, created by Martin Cooper in 1973, there have been significant changes on design, technology and use. It has become a medium that offers a wide variety of integrated functions, such as: an efficient and developed camera, a greater internal storage capacity, and services through the internet, for example, access to social networks or instant messaging with applications like WhatsApp, Skype, Telegram and Facebook.

All this has led to reliance on mobile phone, having users localized and making them being pending of it all the time. Mobile has become a basic and necessary tool in the life of any teenager and adult, using them to feel integrated into society, interacting with their peer group and being alert of the latest developments. It can also cause frustration for not being on fashion and not knowing the latest trends. As shown, the mobile phone is part of everyday life in today's society, without which could hardly survive today. This is demonstrated by studies conducted by Statista Inc. [1] [2]. In Figure 6.1 the results of these studies are shown. As shown in the figure, in recent years there is a steady increase in the number of users with mobile phones and the number of registered users in social networks. This increase was similar in both cases and will continue until 2019.

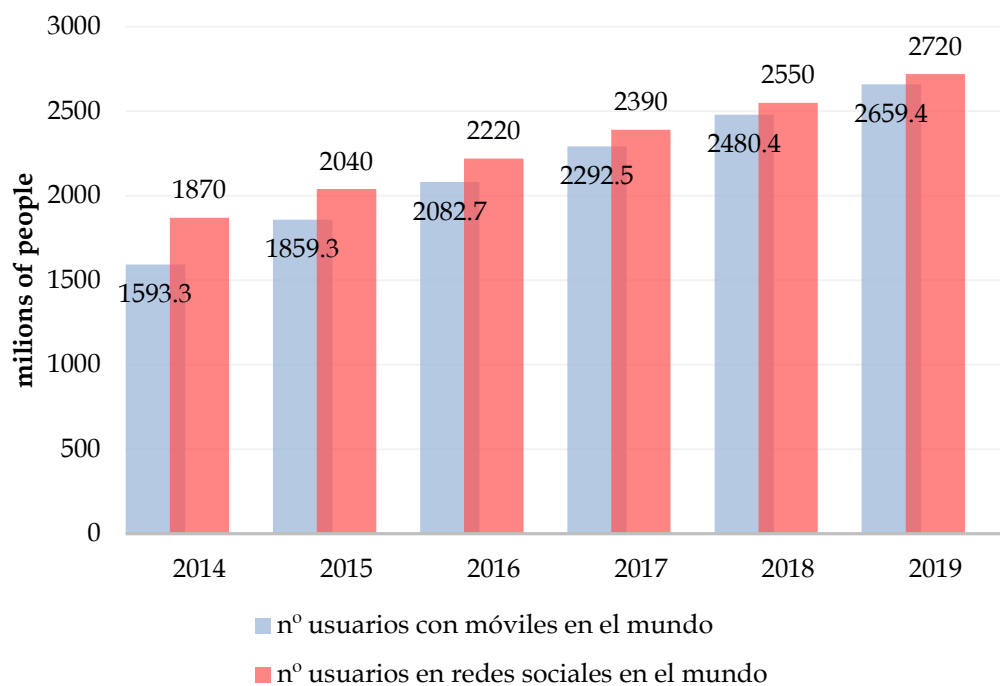


Figure 6.1: Mobile devices and social networks users between 2014 and 2019

As a result of this, it is increasingly common to present as evidence in a trial, WhatsApp messages, SMS, photos, videos or comments shared on social networks such as signs or evidence of a criminal investigation. These tests are incorporated into a case on paper, as documentary evidence, which must be obtained with full guarantees of legality, so that those who provide such evidence should have the right to possession and use of them, as the receiver or the emitter of them.

In the Spanish criminal procedure there is no specific measure on the treatment of evidence of such communications, therefore, is necessary to pay attention to the following articles:

- **Article 26 of the Penal Code:** states that "For the purpose of this code is considered document any material support that expresses or enter data, facts or narratives with probative efficacy or any other legal relevance" [3].
- **Article 726 Criminal Procedure Act:** Establishes that "The Court will examine for itself the books, documents, papers and other material

evidence that may contribute to clarifying the facts or the safest investigation of truth" [4].

The courts usually admit such evidence and incorporate the procedure after performing a verification of them. It has to be recorded the mark and the model of the device from which the evidences come.

Here are some cases in which part of the evidences involve mobile phones:

In the United States, it was investigated in 2015, in the town of Jeanette (Pennsylvania), a murder case that revolved around a Snapchat message, which was used as the main evidence in the case. The case involved a teenager who is facing a charge of first-degree murder of another young man, where the main evidence was a screenshot of a Snapchat message, whose content was a picture of the killer allegedly with the corpse of the victim [5]. A similar case was a person who posted a picture of her beside the body of his partner in Facebook after murder [6]. Finally, in 2014, one person was arrested in Washington D.C. charged with the murder of his partner after a few photos posted on the forum 4chan of the body of the victim [7]. The trail left by the suspects on the Internet was used in these cases as proof of his guilt.

Given the abusive use of the mobile phone and the social networks and the above related evidences for the resolution of a case, it is necessary to conduct an investigation of highly suspicious users that have carried out any inconsistency, since any track, image or threat can help achieve an identification of the source that generated the image or video presented as evidence.

## **7.2. Project Goal**

The main aim of this project is to analyse and understand concepts from the use of mobile phones and social networks to pictures' upload and download from social networks, in order to develop a tool capable of downloading and

applying filters to pictures from the previously mentioned social networks and use them when performing forensic tests. For that reason, the objectives of the project are listed below:

- Understand what a social network is: How they work, what they consist on, how they are used, the different types of them, download and upload of pictures within them, how they process images.
- Develop a tool capable of downloading pictures from a specific album of a user registered in a social network.
- Make some research about the improvements done so far about images for forensic analysis, more in detail, about the identification processes, for a future implementation within the picture download tool.
- Analyse and make some research about the existing tools for downloading images from a social network, and then compare it with the tool created by ourselves, looking for differences and be able of developing a better tool if possible.

Applying the generated tool to a number of users and check whether meet the specifications while the usefulness and reliability of the tool set is checked.

### **7.3. Work Planification**

The Project has been developed in 3 phases: Definition, Implementation and Documentation of the Project. The issues performed in each phase are listed in Table 6.1.

Issue Name	Duration (days)	Beginning	End
<b>Project's Definition</b>	<b>35</b>	<b>02/11/15</b>	<b>18/12/15</b>
• Weekly meetings with the Project's supervisors.	35	02/11/15	18/12/15
• Study of the techniques used in mobile phones for forensic analysis.	15	02/11/15	20/11/15
• Study of the techniques used for forensic analysis applied to mobile's video recorder.	13	23/11/15	09/12/15
• Project's definition	7	10/12/15	18/12/15
<b>Project's Implementation</b>	<b>140</b>	<b>25/01/16</b>	<b>05/08/16</b>
• Requirements specifications	55	25/01/16	08/04/16
• Design	30	11/04/16	20/05/16
• Implementation	29	23/05/16	30/06/16
• Testing operations.	62	20/04/16	14/07/16
• Control.	134	11/01/16	14/07/16
<b>Documentation</b>	<b>200</b>	<b>02/11/15</b>	<b>05/08/16</b>
• Project's documentation generation.	184	02/11/15	14/07/16
• Preparation of the report.	26	01/07/16	05/08/16

Tabla 6.1. Project phases

In the implementation phase is done the development project in the planning phase. In it, the following general activities were carried out: Identification of requirements, design, construction and testing. In addition, monitoring and control project progress were made, in order to expedite the necessary adjustments in each of the activities. Table 6.2 presents the activities in this phase and the duration of them. Figure 6.2 shows the Gantt chart of the project.

Task Name	Time (days)	Start	End
<b>Requirements specifications</b>	<b>55</b>	<b>25/01/16</b>	<b>08/04/16</b>
<ul style="list-style-type: none"> <li>Study of the use of different social networks on mobile devices</li> </ul>	10	25/01/16	05/02/16
<ul style="list-style-type: none"> <li>Analysis of the process of uploading and downloading photos and videos to different social networks</li> </ul>	10	08/02/16	19/02/16
<ul style="list-style-type: none"> <li>Analysis of the changes made to images and videos from social networks</li> </ul>	10	22/02/16	04/03/16
<ul style="list-style-type: none"> <li>Study of image processing techniques</li> </ul>	10	07/03/16	18/03/16
<ul style="list-style-type: none"> <li>Identification of social networks aimed at sharing images and videos</li> </ul>	5	21/03/16	25/03/16
<ul style="list-style-type: none"> <li>Analysis of the functioning of social networks such as Facebook, Flickr, Google +, Tuenti, Twitter and Instagram</li> </ul>	10	28/03/16	08/04/16
<b>Design</b>	<b>30</b>	<b>11/04/16</b>	<b>20/05/16</b>
<ul style="list-style-type: none"> <li>Process Analysis of uploading and manual downloading photos and videos to social networks such as Facebook, Flickr, Google +, Tuenti, Twitter and Instagram.</li> </ul>	7	11/04/16	19/04/16
<ul style="list-style-type: none"> <li>Algorithm design of Facebook downloaded photos</li> </ul>	7	20/04/16	28/04/16
<ul style="list-style-type: none"> <li>Algorithm design of Flickr downloaded photos</li> </ul>	7	29/04/16	09/05/16
<ul style="list-style-type: none"> <li>Design module management downloads</li> </ul>	2	10/05/16	11/05/16
<ul style="list-style-type: none"> <li>Search and analysis of packages or classes that communicate with social networks such as Facebook and Flickr</li> </ul>	7	12/05/16	20/05/16
<b>Implementation</b>	<b>29</b>	<b>23/05/16</b>	<b>30/06/16</b>
<ul style="list-style-type: none"> <li>Study of API of programation on Facebook</li> </ul>	5	23/05/16	27/05/16
<ul style="list-style-type: none"> <li>Implementation of algorithm download of Facebook photos</li> </ul>	7	30/05/16	07/06/16
<ul style="list-style-type: none"> <li>Study of API of programation on Flickr</li> </ul>	5	08/06/16	14/06/16
<ul style="list-style-type: none"> <li>Implementation of flickr algorithm downloads.</li> </ul>	6	15/06/16	22/06/16
<ul style="list-style-type: none"> <li>Implementation of algorithm adjustments identifying the source of acquisition</li> </ul>	6	23/06/16	30/06/16
<b>Tests</b>	<b>62</b>	<b>20/04/16</b>	<b>14/07/16</b>
<ul style="list-style-type: none"> <li>Creation of training dataset with photos and videos taken from mobile devices of different brands and models</li> </ul>	25	20/04/16	24/05/16
<ul style="list-style-type: none"> <li>Algorithm of download Facebook photos</li> </ul>	3	24/06/16	28/06/16
<ul style="list-style-type: none"> <li>Algorithm of download Flickr photos</li> </ul>	3	01/07/16	05/07/16
<ul style="list-style-type: none"> <li>Identification process of photos from mobile devices, downloaded from social networks such as Facebook and Flickr.</li> </ul>	7	06/07/16	14/07/16
<b>Control</b>	<b>134</b>	<b>11/01/16</b>	<b>14/07/16</b>
<ul style="list-style-type: none"> <li>Follow-up weekly meetings</li> </ul>	134	11/01/16	14/07/16

Table 6.2. Activities of the implementation phase of the project

## 7.4. Project Structure

In this chapter and as way of introduction has been written the motivation that has led to the realization of this project and the aim of it.

In Chapter 2, we talk about social networks, use and classification, detailing the most important. Also, was mentioned the addiction that society has nowadays with social networks and mobile phones, as the use of videos and images on social networks has increased significantly.

In Chapter 3 we identified the different techniques for the treatment and the identification of the image source of mobile devices, and the use of sensor imperfections.

In chapter 5 the contribution of this work is presented. It describes the algorithms downloading images of social networks such as Facebook and Flickr, and the identification of the source of images of social networks are listed. In addition, are presented tests to evaluate the performance of the algorithms.

In chapter 6 an analysis for the preparation of the dataset is done, talking about his information and the omission of the study of certain social networks to better understand the need to automate this process and also get a first real experience of analyzing the images metadata.

Finally, in chapter 7 are presented the conclusions that has been reached in this work, and future work that emerges from this study.





## 8. CONCLUSIONS AND FUTURE WORK

---

### 8.1. Conclusions

Most social networks offer users the ability to upload photos to enrich their profile. Some of these social networks are devoted to sharing photos and videos.

Find social networking accounts belonging to a particular person can be very valuable. The task is not trivial, even if details of the identity are known. However, from the perspective of digital forensic investigations, identify the person who generated an image is important in various scenarios. For example, the case of a stolen smartphone that after the robbery, the thief uses it to take pictures and share them on his social networks. Photographs published by the victim in their social networks can be used as samples to correlate the images and identify the thief.

This paper has proposed a system based on the PRNU pattern, combined with clustering techniques to identify the owner of a mobile device. The identification source is performed by analyzing the photos that a user of social networks uploads from the device.

The evaluation of the proposal was made with photos uploaded to Facebook, Flickr and Google +, 3 of the most used social networks nowadays. For this, two algorithms were developed to automatically download pictures of these social networks. The experiments were performed with a set of 3300 photographs of 33 different camera models of mobile devices of 12 manufacturers, which were stored in social networks and then downloaded with the different resolutions available on every social network.

On the results of the experiments it was obtained a success rate of 82.8% in the classification of 11 models of mobile devices from 5 different manufacturers

and an 88.8% success rate in qualifying photographs photos from 5 different models of Samsung manufacturer. The cameras of mobile devices from the same manufacturer should be very similar for much of their products and therefore the extracted noise sensor between different models should resemble. No clutch, it is observed that the correlation value between the PRNU patterns varies enough between models to identify each one separately, even when the models are very similar.

## **8.2. Future Work**

As future work may be noted the following investigation items:

- Expand experimentation to evaluate the use of PRNU pattern in less than 1024 resolutions.
- Evaluate the use of other methods of identifying the source as those based on the characteristics of the image to identify the owner of a photograph
- Expand identification algorithms from stored source to videos on social networks.

## REFERENCIAS

---

- [1] Startista, "Statistics and Market Data on Techonology & Telecommunications", 2012. <https://www.statista.com/markets/418/technology-telecommunications/>
- [2] Startista, "Statistics and Market Data on Social Media & User-Generated Content", 2013. <https://www.statista.com/markets/424/topic/540/social-media-user-generated-content/>
- [3] Jefatura del Estado, "Ley Orgánica 10/1995 de 23 de Noviembre, del Código Penal" BOE No. 281, 1995
- [4] Ministerio de Gracia y Justicia, "Ley de Enjuiciamiento Criminal", BOE A-1992-638, 2015
- [5] B. Stiles, "Jeannette Teen, Charged with Killing Another, Took 'selfie' with Body, Court Papers Say", Trib Live, February 2015. <http://triblive.com/news/westmoreland/7729452-74/police-morton-photo#axzz3Qh0xyVti>
- [6] S. Farberov, "Man Who Shot Dead His Wife Then Posted 'RIP' Picture of Her Corpse on His Facebook Page Played a Gun-Toting Gang Member in TV Series and Has Dreams of Fame", Mail Online, August, 2013. <http://www.dailymail.co.uk/news/article-2387305/Derek-Medina-Florida-man-posts-picture-dead-wife-Jennifer-Alfonso-Facebook.html>
- [7] S. Murdock, "David Kalac Accused of Killing Amber Coplin, Posting Photos of Her Body on 4Chan", The Huffington Post, May, 2014. [http://www.huffingtonpost.com/2014/11/05/david-kalac-4chan\\_n\\_6110716.html](http://www.huffingtonpost.com/2014/11/05/david-kalac-4chan_n_6110716.html)
- [8] A. Martos Rubio, "Redes Sociales", Anaya, Febrero, 2010
- [9] C. C. Aggarwal, "Social Network Analysis", *Data Mining*, pp. 619-661, 2015.

- [10] A. Kuz, M. Falco, R. Giandini, "Análisis de Redes Sociales: un Caso Práctico", *Computación y Sistemas*, Vol. 20, No. 1, 2016.
- [11] S. M. Fernández, F. C. Centellas, "Investigar desde internet: Las redes sociales como abertura al cambio", *Historia y Comunicación Social*, Vol. 18, pp. 663-675, 2013.
- [12] L. Camarero-Cano, "Ensayo Comunidades Tecno Sociales. Evolución de la Comunicación Analógica hacia la Interacción Analógico Digital", *Revista Mediterránea de Comunicación*, 2015.
- [13] I. Ponce, "Redes Sociales-Clasificación de Redes Sociales", Observatorio Tecnológico, 2012.
- [14] M. Zuckerberg, "Facebook", 2004. <https://www.facebook.com/>
- [15] C. Hurley, S. Chen, J. Karim, "Youtube", 2005. <https://www.youtube.com/>
- [16] J. Dorsey, N. Glass, B. Stone, E. Williams, "Twitter", 2006. <https://twitter.com/>
- [17] L. Page, S. Brin, "Google Plus", 2011. <https://plus.google.com/>
- [18] R. Hoffman, A. Blue, K. Guericke, E. Ly, J. Vaillant, "Linkedin", 2002. <https://www.linkedin.com/>
- [19] K. Systrom, M. Krieger, "Instagram", 2010. <https://www.instagram.com/>
- [20] E. Sharp, P. Sciarra, B. Silbermann, "Pinterest", 2008. <https://es.pinterest.com/>
- [21] Ludicorp, "Flickr", 2004. <https://www.flickr.com/>
- [22] Organización Mundial de la Salud, "Informes Técnicos 116 y 117", ONU, 1957
- [23] H. H. Wilmer & J.M. Cheinh. "Mobile Technology Habits: Patterns of Association Among Device Usage, Intertemporal Preference, Impulse Control, and Reward Sensitivity". *Psychonomic Bulletin & Review*, p.1-8, 2016.
- [24] Deloitte, "Deloitte Mobile Consumer Survey, Principales Conclusiones en España", 2015.

- [25] A. Montes-Sotelo, "Adicción a las Redes Sociales", 2016. Benemérita Universidad de Puebla, México, 2016.
- [26] E. Echeburúa, "Adicción a las Nuevas Tecnologías y a las Redes Sociales en Jóvenes: un Nuevo Reto", 2010.
- [27] T. Van Lanh, K. S. Chong, S. Emmanuel, and M. S. Kankanhalli, "A Survey on Digital Camera Image Forensic Methods". In Proceedings of the IEEE International Conference on Multimedia and Expo", pp. 16-19. IEEE, July 2007.
- [28] V. L. L. Thing, K. Y. Ng, and E. C. Chang. "Live Memory Forensics of Mobile Phones", *Digital Investigation*, Vol. 7, pp. S74-S82, August 2010.
- [29] T. Van Lanh, K. S. Chong, S. Emmanuel, and M. S. Kankanhalli, "A Survey on Digital Camera Image Forensic Methods", in Proceedings of the IEEE International Conference on Multimedia and Expo, Beijing, China, July 2007, pp. 16-19.
- [30] M. Boutell and J. Luo, "Beyond Pixels: Exploiting Camera Metadata for Photo Classification", *Pattern Recognition*, Vol. 38, No. 6, pp. 935-946, June 2005.
- [31] D. M. Arenas González, "Análisis Forense de Imágenes de Móviles Mediante el Uso de Metadatos", Universidad Complutense de Madrid, Tesis de Máster 13507, Noviembre 2011.
- [32] K. S. Choi. "Source Camera Identification Using Footprints From Lens Aberration", In Proceedings on Digital Photography II, No. 852 in 6069, pages 60690J-60690J. *SPIE International Society For Optical Engineering*, February 2006.
- [33] S. Bayram, H. T. Sencar, and N. Memon, "Improvements on Source Camera-Model Identification Based on CFA Interpolation", In Working Group 11.9 International Conference on Digital Forensics, pages 24-27. Springer, February 2006.
- [34] O. Celiktutan, I. Avcibas, B. Sankur, N. P. Ayerden, and C. Capar. "Source Cell-Phone Identification", In Proceedings of the IEEE 14th Signal Processing and Communications Applications, pages 1-3. IEEE, April 2006.

- [35] Y. Long and Y. Huang. "Image Based Source Camera Identification Using Demosaicking", In Proceedings of the IEEE 8th Workshop on Multimedia Signal Processing, pages 419-424. IEEE, October 2006.
- [36] S. Bayram, H. T. Sencar, and N. Memon. "Classification of Digital Camera-Models Based on Demosaicing Artifacts", Digital Investigation, Vol. 5, No. 1, pp. 49-59, September 2008.
- [37] M. J. Tsai, C. L. Lai, and J. Liu. "Camera/Mobile Phone Source Identification for Digital Forensics", In Proceedings of the International Conference on Acoustics Speech and Signal Processing, pp. 221-224. IEEE, April 2007.
- [38] C. Mckay, A. Swaminathan, H. Gou, and M. Wu. "Image Acquisition Forensics: Forensic Analysis to Identify Imaging Source", In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, International Conference on Acoustics Speech and Signal Processing (ICASSP), pages 1657-1660. IEEE, June 2008.
- [39] F. J. Meng, X. W. Kong, and X. G. You. "Source Camera Identification Based on Image Bi-Coherence and Wavelet Features", In Proceedings of the Fourth Annual IFIP WG 11.9 International Conference on Digital Forensics, Kyoto, Japan, January 2008.
- [40] B. Wang, Y. Guo, X. Kong, and F. Meng. "Source Camera Identification Forensics Based on Wavelet Features", In Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pages 702-705. IEEE Computer Society, September 2009.
- [41] Z. J. Geradts, J. Bijhold, M. Kieft, K. Kurosawa, K. Kuroki, and N. Saitoh. "Methods for Identification of Images Acquired with Digital Cameras", In Proceedings on Enabling Technologies for Law Enforcement and Security, volume 4232, pages 505-512. SPIE-International Society for Optical Engineering, February 2001.
- [42] J. Lukas, J. Fridrich, and M. Goljan. "Digital Camera Identification from Sensor Pattern Noise", IEEE Transactions on Information Forensics and Security, Vol. 1, No. 2, pp. 205-214, June 2006.

- [43] F. D. O. Costa, M. Eckmann, W. J. Scheirer, and A. Rocha, "Open Set Source Camera Attribution", In Proceedings of the 25th Conference on Graphics, Patterns and Images, pp. 71-78. IEEE, August 2012.
- [44] J. Fridrich, "Digital Image Forensics", *IEEE Signal Processing Magazine*, Vol. 26, No. 2, pp. 26-37, March 2009.
- [45] C.-T. Li, "Unsupervised Classification of Digital Images Using Enhanced Sensor Pattern Noise", in Proceedings of the IEEE International Symposium on Circuits and Systems. IEEE, May 2010, pp. 3429-3432.
- [46] R. Caldelli, I. Amerini, F. Picchioni, and M. Innocenti, "Fast Image Clustering of Unknown Source Images", in Proceedings of the IEEE International Workshop on Information Forensics and Security. IEEE, December 2010, pp. 1-5.
- [47] L. Rokach, "A Survey of Clustering Algorithms", in Data Mining and Knowledge Discovery Handbook, O. Maimon and L. Rokach, Eds. Springer US, 2010, pp. 269-298.
- [48] J. Rosales Corripio, D. M. Arenas González, A. L. Sandoval Orozco, L. J. García Villalba, J. C. Hernandez-Castro, and S. J. Gibson, "Source Smartphone Identification Using Sensor Pattern Noise and Wavelet Transform", in Proceedings of the 5th International Conference on Imaging for Crime Detection and Prevention, London, UK, December 2013, pp. 1-6.
- [49] ThoughtWorks, "Selenium Automates Browsers", <http://www.seleniumhq.org/>